

ESSAYS ON SHOPPING DYNAMICS IN CUSTOMER BASE ANALYSIS

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Chang Hee Park

August 2012

© 2012 Chang Hee Park
ALL RIGHTS RESERVED

ESSAYS ON SHOPPING DYNAMICS IN CUSTOMER BASE ANALYSIS

Chang Hee Park, Ph.D.

Cornell University 2012

Given the emerging concept of a customer-centric approach to marketing, customer relationship management (CRM) has seen increased attention. Among essential tools to implement CRM is customer base analysis which seeks to understand and predict transaction patterns of individual customers. This dissertation, composed of three essays, studies the dynamics of shopping behavior in customer base analysis and its implications for CRM.

The first essay provides an overview of modeling approaches for customer base analysis, reviews relevant research in the marketing literature, and identifies an agenda of areas that are in need for further research.

The second essay proposes a modeling framework for multi-category customer lifetime value (CLV) analysis in a non-contractual setting. To this end, we model customers' arrival process, purchase incidence and amount decisions across categories, and latent defection in an integrated framework. The proposed framework makes use of a latent space model that parsimoniously captures various dynamics of multi-category shopping behavior arising from the interplay between purchase timing and choice across categories. Using category-level transaction data from a leading beauty care company, we show that the proposed model offers excellent fit and performance in predicting customer purchase patterns across categories. Our model allows one to quantify the contribution of individual categories to CLV and assess the relationship between shopping basket choice and CLV.

The third essay examines shopping behavior of online customers. We develop a model that captures the clustered visit patterns of online customers and predicts how a series of store visits lead to a purchase. Our model is based on the notion that the arrival process of customer visits consists of multiple visit clusters with relatively short intervisit times within a cluster and a longer intervisit time between clusters. Because the start and the end of each visit cluster are unobserved, we employ a changepoint modeling framework and statistically infer the cluster formation on the basis of customer visit patterns through data augmentation in Bayesian approach. In our empirical analysis using data from a major e-commerce site, we find strong empirical evidence of lumpy shopping patterns by online customers with significant heterogeneity in the extent of the lumpiness. As part of our substantive contribution, we show that taking into account the clustered visit patterns can significantly improve the model performance in predicting purchase conversions across store visits.

BIOGRAPHICAL SKETCH

Chang Hee Park was born in South Korea in December 1979. In March 1997, he matriculated at Korea Advanced Institute of Science and Technology (KAIST), where he majored in Industrial Engineering. During his undergraduate studies at KAIST, his interest was focussed on the applications of mathematical and engineering methodologies to business decision-making problems. He graduated from KAIST with a B.S. degree in February 2001, and worked as a developer of management information systems for three years. Chang Hee came to the United States in August 2004 to attend a Master's program in Management Science and Engineering at Stanford University, where he further enhanced his quantitative skills. He was awarded an M.S. degree in June 2006. His intellectual interest in quantitative marketing models led him to join the Ph.D. program at the Johnson School of Management at Cornell University in August 2006. At Cornell University, Chang Hee performed research on the dynamic aspects of customer shopping behavior and their implications for customer relationship management. He received a Ph.D. degree in Management in August 2012.

This dissertation is dedicated to my family

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor, Professor Young-Hoon Park, for the enormous amount of support and encouragement he has provided throughout my Ph.D. study. He has truly exemplified a level of scholarship to which I aspire, and constantly inspired me by sharing his insights, passions, and philosophies. I am deeply indebted to him and wholeheartedly appreciate his guidance and concern for my academic advancement.

I have greatly benefited from the interactions with my committee members, Professor Vithala R. Rao, Professor Sachin Gupta, and Professor Vishal Gaur, who have played critical roles in my academic development. They have always given me constructive criticisms and suggestions on my work. My special thanks go to Professor Vithala R. Rao for his genuine support for doctoral students in Marketing at the Johnson School.

I am profoundly grateful to my parents and parents-in-law for their endless love and support. They have been an eternal source of warmth and wisdom in my life. They have guided me to become the person I am today, and their dedication made the completion of this degree possible.

Finally, my deepest gratitude goes to my wife and soul mate, Jiyoung Lee. The numerous sacrifices she has made throughout my Ph.D. study have allowed me to pursue my aspirations and reach important milestones in my professional career. She has always believed in me and immensely supported me. This dissertation would not have been possible without her. Our lovely daughters, Sieun and Gaeun, have been a real blessing in our life ever since they came to us. I dedicate this dissertation to my family.

TABLE OF CONTENTS

Biographical Sketch	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vi
List of Tables	viii
List of Figures	ix
1 Marketing Models for Customer Base Analysis	1
1.1 Introduction	1
1.2 Two Modeling Approaches	3
1.2.1 Regression-Type Models	3
1.2.2 Probability Models	5
1.3 Literature Review	6
1.3.1 Allowing for Nonstationarity	7
1.3.2 Modeling Interdependent Processes	8
1.3.3 Incorporating Customer Defection	9
1.4 Future Research in Customer Base Analysis	12
1.5 Contribution of This Dissertation	15
2 Valuing Customers Category by Category: An Integrated Model for Multi-category Customer Lifetime Value Analysis	22
2.1 Introduction	22
2.2 Previous Research	26
2.3 Model Development	29
2.3.1 Timing Model	29
2.3.2 Multi-category Choice Model	35
2.3.3 Purchase Amount Model	37
2.3.4 Customer Defection	38
2.3.5 Customer Heterogeneity and Correlation Structure	40
2.3.6 Computational Approach	41
2.4 Data	42
2.5 An Empirical Application	46
2.5.1 Selecting the Number of Dimensions in the Latent Space	46
2.5.2 Model Comparison	48
2.5.3 Model Fit	51
2.5.4 Parameter Inferences	57
2.5.5 Predicting CLV at the Category Level	65
2.5.6 Linking Shopping Basket Choice to CLV	69
2.6 Conclusions and Future Research	73

3	Modeling Online Visitation and Conversion Dynamics	81
3.1	Introduction	81
3.2	Model	87
3.2.1	Model Structure	87
3.2.2	Purchase Model	88
3.2.3	Visit Model	90
3.2.4	Computational Approach	98
3.3	Data	102
3.4	An Empirical Application	104
3.4.1	Model Specification	104
3.4.2	Model Fit	106
3.4.3	Model Results	108
3.5	Conclusions and Future Research	120

LIST OF TABLES

2.1	Number of Parameters Required to Consider the Effect of Shopping Basket Choice on Arrival Rate	35
2.2	Descriptive Statistics of the Data	43
2.3	Shopping Basket and Interarrival Time	45
2.4	Selection of the Number of Dimensions in the Latent Space	47
2.5	Model Comparison	50
2.6	Parameter Estimates of the Timing Model: Baseline Arrival Rate and Customer Defection	57
2.7	Parameter Estimates of the Timing Model: Coordinates of Categories in a Three-Dimensional Latent Space	58
2.8	Parameter Estimates of the Choice Model: Category-Specific Effects	59
2.9	Parameter Estimates of the Choice Model: Cross-Category Effects	60
2.10	Parameter Estimates of the Choice Model: Lagged Effects	61
2.11	Parameter Estimates of the Purchase Amount Model	62
2.12	Estimated Σ	64
2.13	Correlation of CLV between Categories	68
2.14	Comparisons of CLV between Customer Groups	71
3.1	Descriptive Statistics of the Data	103
3.2	Data Summary on Reference Sites	104
3.3	Model Comparisons	108
3.4	Parameter Estimates of the Purchase Model	109
3.5	Conversion Rates for the Sequence of Reference Sites in Clusters with Two Visits	113
3.6	Parameter Estimates of the Visit Model	115
3.7	Estimated Σ	115
3.8	Customer-Specific Statistics of Latent Visit Clusters	117
3.9	Comparison of Two Customers	119

LIST OF FIGURES

1.1	Purchase Patterns of a Customer	2
2.1	Two-Category Representation in a Two-Dimensional Space . . .	32
2.2	Model Fit on Interarrival Time after Purchasing a Shopping Basket	52
2.3	Model Fit on Cumulative Purchases of a Shopping Basket	54
2.4	Model Fit on Mean Purchase Amount	55
2.5	Model Fit on Cumulative Purchase Amount	56
2.6	Prediction of Three-Year Customer Lifetime Value	67
2.7	Contribution of Individual Categories to CLV	69
3.1	Lumpy Shopping Patterns of an Online Customer	83
3.2	Conversion Rate in a Visit Cluster	111

CHAPTER 1

MARKETING MODELS FOR CUSTOMER BASE ANALYSIS

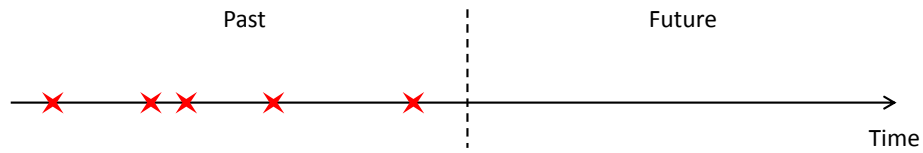
1.1 Introduction

Over the past decade, customer relationship management (CRM) has been one of the most prominent aspects of marketing strategies and activities by firms (Reinartz, Krafft, and Hoyer 2004, Payne and Frow 2005). Developing and managing long-term relationships with customers often rely on the prediction of their future purchase behavior (Gupta et al., 2006). Accordingly, customer base analysis has seen increased attention among both practitioners and researchers.

Consider a customer who made five transactions with a firm in the past (or in the period of transaction database). Figure 1.1 illustrates her purchase patterns from a timing perspective. Given the customer's purchase history, marketers would be interested in making forecasts about her future behavior. The main purpose of customer base analysis is to address such predictive questions as (1) how many customers are currently active, (2) how that number will change over time, (3) which customers are more likely to remain active in a future time period, and (4) how many purchases the customers will make before becoming inactive, and what their purchase patterns will look like (Schmittlein and Peterson 1994).

The benefits of these projections can range from aggregate sales trajectories (e.g., the total number of purchases by existing customer base for the next 6 months) to individual-level conditional expectations (e.g., the number of transactions by a particular customer conditional on her past behavior) (Fader and

Figure 1.1: Purchase Patterns of a Customer



Hardie 2009). In particular, the understanding of customer base is of great help for marketers in tailoring marketing offers and developing relationships with customers in order to increase profitability at the customer level. For example, marketers can customize promotional messages based on the state of customers (e.g., active versus inactive) and decide how much should be spent to manage customers by taking into account the expected future cash flows from individuals (Kumar 2008). Other application areas include customer segmentation based on predicted outcome measures and the evaluation of communication channels/marketing contacts (Schmittlein and Peterson 1994). The concrete benefits of customer base analysis, coupled with the growing availability of customer-level transaction data, have motivated ample research on it. The purpose of this chapter is to provide an overview of relevant studies in the marketing literature and discuss potential areas for future research.

The remainder of this chapter is organized as follows. Section 1.2 introduces key modeling approaches of predicting customer behavior. In section 1.3, we review prior research. Section 1.4 identifies potential research topics for future work. We conclude by discussing the contribution of this dissertation to the stream of research.

1.2 Two Modeling Approaches

In this section, we discuss modeling approaches commonly used by researchers to predict purchase behavior. The existing models can be broadly classified into two categories of regression-type models and probability models, depending on whether they employ the concept of customers' latent characteristics in describing observed behavior and linking it to future behavior. Below, we introduce each class of models and compare/contrast them with each other.

1.2.1 Regression-Type Models

The standard approach taken by researchers who attempt to make forecasts about customer behavior is to use regression-type models (also called scoring models) (e.g., Bolton 1998, Malthouse and Blattberg 2005). In regression models, a dependent variable (e.g., the number of future transactions) is specified as a function of several explanatory variables constructed based on a customer's demographic information and behavioral characteristics. RFM measures – recency (the time of most recent purchase), frequency (the number of past purchases), and monetary value (the average purchase amount per transaction) – serve as well-established metrics that summarize a customer's past purchase behavior (Hughes 2005, Gupta et al. 2006). To calibrate a regression-type model, researchers generally split the transaction data into pairs of consecutive time periods: Data from the first period of each pair (e.g., period 1) are used to generate the explanatory variables, whereas data from the second period of the pair (e.g., period 2) create the dependent variable.

While the regression-type models are intuitive and straightforward in its implementation, there are several limitations with them. (See Fader, Hardie, and Lee (2006) for more detailed discussion.) Most importantly, regression models make prediction for the next period only. For example, having calibrated the regression model on data from period 1 to 3, we can predict buying behavior in period 4. However, it is not clear how we can forecast behavior in period 5 because the values of behavioral predictor variables in the model are not available in period 4. One may attempt to model all of these covariates and use their predicted values in period 4 to forecast buying behavior in period 5. However, as we move further away from the data period, it would become increasingly difficult to obtain reliable results because, with multiple interacting variables, a small change in one variable can lead to dramatic differences in the forecasts after a few periods.

Another problem with regression models is that behavioral predictor variables are unstable indicators of true behavior. For example, RFM variables vary over time. So, different ways of slicing the data (e.g., weekly, monthly) can result in different values of the variables and thus provide different forecasts and inferences. In addition, because the dependent variable is derived from customers' behavior, it is more reasonable to treat it as a random variable rather than a fixed number. However, such stochastic nature of the variable is not explicitly considered in the standard regression model.

1.2.2 Probability Models

Recognizing the limitations of regression-type models, researchers have taken an alternative modeling approach. In this way, a customer's observed behavior is regarded as the outcome of an underlying stochastic process governed by her latent behavioral characteristics. A mathematical model is developed to specify the relationship between observed behavior and latent characteristics (i.e., observed behavior = $f(\theta)$, where θ denotes latent behavioral characteristics). In many cases, this is done by employing a combination of probability distributions which characterize observed behavior with the assumption that their parameters capture latent characteristics. We then estimate the model parameters (i.e., infer a customer's latent characteristics) given observed behavior (i.e., $\hat{\theta} = g(\text{past behavior})$). This allows us to make prediction about future behavior by plugging the estimated parameters into the model (i.e., future behavior = $f(\hat{\theta})$). A class of models under this approach is often referred to as probability models.

The two-step approach in probability models allows us to avoid the problems with single-step regression-type models (i.e., future behavior = $f(\text{past behavior})$). In particular, using a probability model, we can predict future behavior over time periods of any length, because the predictions are made based on the inferred behavioral characteristics, not directly from past behavior. Also, the quality of inferences using a probability model are much less influenced by the way of slicing data. Given these benefits, there has been a strong tradition of probability models in marketing. The following section reviews this stream of research.

1.3 Literature Review

In this section, we review probability models for customer base analysis in the marketing literature. We begin with the discussion of stationary models. We then discuss their notable extensions, classified into three categories: (1) allowing for nonstationarity in customer purchase patterns, (2) modeling multiple interdependent processes, and (3) incorporating customer defection into the purchase process.

Predicting customers' purchase patterns has been of great interest to marketing researchers. One of the widely used modeling frameworks is the negative binomial distribution (NBD) model which characterizes the interarrival times between a customer's consecutive transactions by an exponential distribution (or equivalently assumes that the number of purchases made by an individual in a given time period follows a Poisson distribution) and accounts for customer heterogeneity in the buying rate by a gamma distribution (e.g., Ehrenberg 1988, Wellan and Ehrenberg 1988, Morrison and Schmittlein 1988). Many researchers have showed that the NBD model offers an excellent empirical performance in various contexts. However, questioning the unrealistic memoryless property of an exponential distribution, others have used an Erlang-2 or Weibull distribution as an alternative to the exponential distribution (e.g., Gupta 1988, Wheat and Morrison 1990).

In the NBD model (or other comparable timing models using an Erlang-2 or Weibull distribution), a customer's purchase process is assumed to be stationary, meaning the buying rate is time-invariant. The model appears to capture the transaction patterns quite well when the assumption of stationarity is satis-

fied in data. However, if the underlying process is nonstationary (i.e., the buying rate is time-varying), the projections of purchase patterns using the NBD model are likely to be biased.

1.3.1 Allowing for Nonstationarity

Recognizing the limitation of the NBD model, several researchers have extended the model to take into account nonstationarity. A common approach of allowing for nonstationarity is to include time-varying covariates in the model so that the buying rate can vary over time depending on the covariates. Information captured by time-varying covariates includes firms' marketing actions, lagged variables (e.g., lagged interpurchase times), seasonal factors, and macroeconomic variables (for long-term studies), just to name a few. Gupta (1991) makes an important contribution in this area.

The NBD-based model by Gupta (1991) is not the only way to consider the effects of covariates on customer purchase patterns and another popular modeling approach is the proportional hazard model, first proposed by Cox (1972). In the proportional hazard model, the hazard rate, the event rate at time t conditional on the non-occurrence of the event until time t , is decomposed into two multiplicative components: the baseline hazard rate with a prespecified distribution (e.g., an exponential distribution) and the covariate function. Several researchers have employed or extended the proportional hazard model to study the influence of marketing mix variables on customers' purchase timing decisions (e.g., Jain and Vilcassim 1991, Wedel et al. 1995, Seetharaman and Chintagunta 2003, Telang, Boatwright, and Mukhopadhyay 2004).

In some cases, a customer's buying behavior evolves over time for unobservable reasons such as her learning about products and preference changes. In these situations, it is difficult to not only define covariates that explain the resulting dynamics but also collect relevant information to construct the variables. To deal with this, researchers have developed dynamic models which systematically update the (baseline) buying rate based on observed behavior. For example, Moe and Fader (2004) extend the NBD model by allowing customers' visit rate to continuously change from one visit to the next in an online shopping context. Fader, Hardie, and Huang (2004) develop a changepoint model that detects underlying shifts in customers' consumption patterns of a new product, as a result of their acquisition of experience with a new product. In a similar context, Schweidel and Fader (2009) study the scenario that the underlying purchase process itself evolves from an initial state characterized by an exponential timing process to a steady state of Erlang-2 timing process. These models can be regarded as continuous-state or discrete-state hidden Markov models, which have been applied to model a wide range of latent changes in customer behavior (e.g., Montgomery et al. 2004, Netzer, Lattin, and Srinivasan 2008).

1.3.2 Modeling Interdependent Processes

The increasing availability of customers' shopping data across multiple product categories/firms has motivated researchers to develop an integrated framework to simultaneously model multiple interdependent purchase processes. A common approach for this purpose is to use a multivariate distribution that ties together timing processes of interest. Chintagunta and Haldar (1998) are among the first who bring this idea to the marketing literature. Using the Farlie-

Gumbel-Morgenstern (FGM) family of distributions (Johnson and Kotz 1975), they develop a bivariate proportional hazard model to study the relationship in purchase timing of two related categories (e.g., pasta and pasta sauce).

While the FGM approach offers a reasonably straightforward way of capturing dependence between distributions, it has limitations from an inferential standpoint because it does not generally yield marginal distributions that match the functional forms of the designated univariate densities (Park and Fader 2004). The Sarmanov family of distributions (Sarmanov 1966) is a class of multivariate distributions that avoid this undesirable property while offering more flexibility in capturing the correlation of distributions. Using the Sarmanov distribution, Park and Fader (2004) develop a NBD-based timing model in which they examine online customers' visit behavior across websites. Schweidel, Fader, and Bradlow (2008) apply the Sarmanov distribution to the proportional hazard model in order to examine the relationship between customer acquisition and retention processes. These models offer superior fits and predictive performance compared to their counterpart independent models. However, perhaps due to the accompanying complexities of using the multivariate distributions, most research focuses on the bivariate applications of the models.

1.3.3 Incorporating Customer Defection

In most business contexts, the customer-firm relationship is not everlasting and customers terminate their relationship with the firm at some point. Accordingly, failing to consider their defection can result in erroneous inferences on purchase behavior (e.g., the underestimation of buying rates while active). In

this regard, another important extension of the NBD-based models is to allow for the possibility of customer attrition. This addition offers a significant managerial benefit to marketers because distinguishing between customers who are active and those who have become inactive is central to various managerial decisions such as customer targeting, resource allocation, and customer valuation (e.g., Reinartz and Kumar 2000).

The contexts of modeling customer defection are generally divided into contractual and noncontractual settings. The defining characteristic of a contractual setting is that the time at which a customer becomes inactive is observed by the firm (e.g., magazine subscription, insurance policy purchase). In contrast, a noncontractual setting is one where the firm does not observe customer defection (e.g., grocery shopping, hotel stays). Thus, the firm cannot certainly distinguish between a customer who has ended her relationship with the firm and one who is merely in the middle of a long hiatus between transactions.

Noncontractual contexts are very common in practice, yet predicting customer purchase patterns in these contexts is challenging because the firm never knows whether a customer has defected or not (Bell et al. 2002, Singh, Borle, and Jain 2009). Schmittlein, Morrison, and Colombo (1987) propose the Pareto/NBD model which explicitly addresses this problem. The Pareto/NBD model specifies a customer's relationship with the firm in two phases: she is active (or "alive") for an unobserved time period, and then becomes permanently inactive (or "dead"). While the customer is active, her purchase behavior is characterized by the NBD model. The customer's unobserved lifetime is modeled by an exponential distribution, and heterogeneity in dropout rates across customers is considered through a gamma distribution. By coupling the two timing pro-

cesses for customer arrival and latent attrition, this “buy ‘til you die” model allows us to assess a customer’s lifetime in a probabilistic manner.

The Pareto/NBD model has been extended in a number of directions. Schmittlein and Peterson (1994), Fader, Hardie, and Lee (2005a), and Singh, Borle, and Jain (2009) augment the model with a submodel of purchase amount to predict customer lifetime value (CLV). Abe (2009) incorporates customer-specific covariates into the specification of buying and drop rates. Fader, Hardie, and Lee (2005b) develop the BG/NBD model to avoid the computation difficulties in estimating the Pareto/NBD model. The assumptions on the interarrival times in the BG/NBD model are the same as those in the Pareto/NBD model. However, unlike the Pareto/NBD model, the BG/NBD model assumes that a customer can become inactive after any transaction with some probability and heterogeneity in dropout probabilities across customers is considered through a beta distribution. The BG/NBD model is easier to implement but still provides a similar forecasting performance with the Pareto/NBD model. The Pareto/NBD model is a continuous-time model in which customer transactions and defections can occur at any point in time. Fader, Hardie, and Shang (2010) propose the BG/BB model as a discrete-time analog of the Pareto/NBD model.

Predicting customer transaction patterns in a contractual setting is relatively simple compared to a noncontractual setting, because we do not face the problem of distinguishing between a customer who has ended her relationship with the firm and one who is in a long hiatus between transactions. Nevertheless, customer base analysis in a contractual setting requires a well-defined predictive model to answer questions such as: how likely customers will stay with the firm in a next period, and what are factors driving or indicating the risk of

churning (Fader and Hardie 2009).

The shifted-beta-geometric model has been widely used for analysis in a discrete-time setting (e.g., Weinberg and Gladen 1986, Fader and Hardie 2007). The model assumes that a customer remains with some retention probability at the end of each contract period, and the retention probability follows a beta distribution. A common modeling approach with continuous-time duration data is to use the exponential-gamma model which characterizes interarrival times between renewals by an exponential distribution and accounts for customer heterogeneity in the arrival rate by a gamma distribution (e.g., Morrison and Schmittlein 1980, Dekimpe and Morrison 1991, Hardie, Fader, and Wisniewski 1998). Alternatively, Braun and Schweidel (2011) employ the proportional hazard model and combine it with the competing risk model to link the different reasons for which customers churn to their value to the firm.

1.4 Future Research in Customer Base Analysis

We have reviewed probability models for customer base analysis in the marketing literature. The common thread that underlies this class of models is the approach taken in model development. Unlike the regression-type models which build a direct functional link from past behavior to future behavior, probability models view observed behavior as the realization of an underlying stochastic process governed by latent behavioral characteristics. The focus of much of this pioneering work has been on building a flexible yet parsimonious framework in its implementation.

Based on the state of existing modeling tools as reflected in the literature

and the needs of marketing managers to make the most of their customer-level database, we suggest a number of avenues for future research:

- **Incorporating Marketing Decision Variables**

While several researchers have developed timing models which account for the impacts of a firm's marketing activities on customer purchase patterns (e.g., Gupta 1991, Fader, Hardie, and Huang 2004), extant "buy 'til you die" models lack such considerations. The importance of incorporating marketing mix variables cannot be understated in a noncontractual setting, because they can affect customer defection as well as buying rates while active. When marketing mix variables are included in the model, the prediction of future behavior is not straightforward, because the values of the covariates are only available in the data period and we have no information about the firm's marketing actions in a future time period. One way to deal with this problem is to model not only customer behavior given the firm's marketing actions but also the firm's marketing decisions given customer behavior in an integrated framework. By doing so, we can generate interdependent sequences of customer transactions and marketing activities beyond the data period, which can be used for prediction and inference purposes.

- **Forecasting Future Behavior at Early Relationship Stages**

As a key byproduct of customer base analysis, CLV is a metric to measure the relative attractiveness of individual customers. Computing CLV relies on the prediction of individuals' future purchase patterns based on their past behavior. This means that for reasonably accurate prediction of CLV, we need to wait for a while until a certain amount of transaction information of individual customers is accumulated. Yet, predicting CLV at very

early stages is of particular importance to firms because it can significantly help shape and direct a nascent relationship with customers. For example, if customers with a high CLV can be identified even before they have proven to be high-value customers, firms can take preemptive actions to retain the customers. While the determinants of CLV has been studied in the literature (e.g., Thomas 2001, Hitt and Frei 2002), little research attempts to link CLV to early-stage behavioral characteristics of customers.

- **Understanding Dynamics of Customer Behavior**

Given the increasing variety of customer-level transaction data, there would be opportunities to gain new insights about how unique features of a firm's business practices may affect customer behavior over time. One example is to incorporate the influence of loyalty marketing (e.g., rewards programs, cash-back offers) into customer base analysis. By offering benefits based on cumulative purchasing over time, loyalty programs tend to encourage consumers to shift from myopic or single-period decision making to dynamic or multiple-period decision making (e.g., Lewis 2004). This in turn would affect the customers' consumption patterns and lifetime values. Such aspects have not been investigated in the existing literature.

- **Capturing Network Effects**

Most research in customer base analysis assumes that purchasing patterns are independent across customers. However, in many situations, the effects of social networks and word-of-mouth between customers may be strong, and thus accounting for them could help better predict customer retention and attrition (e.g., local people's restaurant choices, visiting an e-commerce site about which customers actively communicate online). Prior research also provides empirical evidence that an individual's preference

for a product is influenced by the preferences of others (e.g., Yang and Allenby 2003, Ansari, Koenigsberg, and Stahl 2011). It would be fruitful to develop a framework which accounts for evolving networks and takes into account their dynamic impacts on customer base.

- **Dealing with Aggregated Data**

In some cases, we do not have access to customer-level transaction data. Instead, we only have a series of cross-sectional summaries across customers. Without individual-level information, most of the existing “buy ‘til you die” models are not applicable, but marketing managers may still seek to understand their customer base at least at the aggregate level (e.g., the number of active customers in a future time period, the average number of transactions made by them before becoming inactive). Fader, Hardie, and Jerath (2007) show how the Pareto/NBD model can be estimated in such cases. This area of research is still in its infancy and there is a growing need for parsimonious models applicable with coarse data.

1.5 Contribution of This Dissertation

The remainder of this dissertation consists of two essays which incorporate the dynamics of shopping behavior into customer base analysis in a noncontractual setting, and study their implications for CRM. We conclude this chapter by summarizing the essays and discussing their contribution to the literature.

The first essay proposes a modeling framework to extend customer base/lifetime value analysis to a multi-category transactional context. We model customers’ arrival process, purchase incidence and amount decisions

across categories, and latent defection in an integrated framework, allowing for the dynamics of shopping behavior due to the interdependence between category purchases and interarrival times. We also account for the interdependent choices of multiple categories, the correlation of repeatedly observed choice outcomes both within and across categories, and latent customer defections. One of the major complications in modeling the association of shopping baskets and the arrival process is that the number of possible compositions of shopping baskets (and thus the complexity of the model) increases exponentially as the number of categories increases. To avoid this “curse of dimensionality,” we propose a novel modeling approach using a latent space of product categories. Our approach is easily extendable to multi-category cases and yields a parsimonious model for customer base analysis with multiple categories.

We apply our model to category-level transactional data from a leading beauty care company in Korea. We empirically assess the need to include the different modeling components and find support for the inclusion of each. As a result, the proposed model offers excellent performance of predicting customer purchase patterns of multiple categories. Our model allows one to quantify the contribution of individual categories to CLV and assess the relationship between shopping basket choice and CLV.

The second essay examines the shopping patterns of online customers at an e-commerce site and examines how a series of store visits lead to a purchase. Toward this end, we develop a dynamic timing model that explicitly captures the lumpy visit patterns of online shoppers. Our model is based on the notion that the arrival process of customer visits consists of multiple visit clusters with a relatively high visit rate within a cluster and a lower visit rate between

clusters. Because the start and the end of each visit cluster are unobserved, we employ a changepoint modeling framework and statistically infer the cluster formation through data augmentation in Bayesian approach. A key benefit of our dynamic model is its ability to infer the formation of latent visit clusters in the arrival process of customer visits, which offers a set of novel inferences about the patterns underlying online shopping behavior, including (1) the number of visits per cluster, (2) the intervisit time within a cluster, (3) the time length of a cluster, (4) the number of visit clusters in a given time period, and (5) the intervisit time between clusters, at the individual customer level.

In our empirical analysis using data from a major e-commerce site in the United States, we find strong empirical evidence of lumpy shopping patterns by online customers with significant heterogeneity in the extent of the lumpiness. As part of our substantive contribution, we highlight how the cluster-based inferences of store visit patterns can be leveraged to examine conversion behavior of online customers. We find that a customer’s purchase likelihood varies across visits depending on how many visits the customer has made previously within a visit cluster. As a result, we show that taking into account the clustered visit patterns can significantly improve the model performance in predicting purchase conversions.

References

- Abe, Makoto (2009), "'Counting Your Customers' One by One: A Hierarchical Bayes Extension to the Pareto/NBD Model," *Marketing Science*, 28 (3), 541-553.
- Ansari, Asim, Oded Koenigsberg, and Florian Stahl (2011), "Modeling Multiple Relationships in Social Networks," *Journal of Marketing Research*, 48 (4), 713-728.
- Bell, David, John Deighton, Werner J. Reinartz, Roland T. Rust, and Gordon Swartz (2002) "Seven Barriers to Customer Equity Management," *Journal of Service Research*, 5 (1), 77-86.
- Bolton, Ruth N. (1998), "A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction," *Marketing Science*, 17 (1), 45-65.
- Braun, Michael and David A. Schweidel (2011), "Modeling Customer Lifetimes with Multiple Causes of Churn," *Marketing Science*, Forthcoming.
- Chintagunta, Pradeep K. and Sudeep Halder (1998), "Investigating Purchase Timing Behavior in Two Related Product Categories," *Journal of Marketing Research*, 35 (1), 43-53.
- Cox, David R (1972), "Regression Models and Life-Tables," *Journal of the Royal Statistical Society, Series B*, 34 (2), 187-220.
- Dekimpe, Marnik G. and Donald G. Morrison (1991), "A Modeling Framework for Analyzing Retail Store Durations," *Journal of Retailing*, 67 (1), 68-92.
- Ehrenberg, A. S. C. (1988), *Repeat-Buying: Facts, Theory and Applications*, 2nd edn., London: Griffin; New York: Oxford University Press.
- Fader, Peter S. and Bruce G. S. Hardie (2007), "How to Project Customer Retention," *Journal of Interactive Marketing*, 21 (1), 76-90.
- Fader, Peter S. and Bruce G. S. Hardie (2009), "Probability Models for Customer-Base Analysis," *Journal of Interactive Marketing*, 23 (1), 61-69.
- Fader, Peter S., Bruce G. S. Hardie, and Chun-Yao Huang (2004), "A Dynamic Change-point Model for New Product Sales Forecasting," *Marketing Science*, 23 (3), 280-303.
- Fader, Peter S., Bruce G. S. Hardie, and Kinshuk Jerath (2007), "Estimating CLV Using Aggregated Data: The Tuscan Lifestyles Case Revisited," *Journal of Interactive Marketing*, 21 (3), 55-71.

- Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005a), "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis," *Journal of Marketing Research*, 42 (4), 415-430.
- Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005b), "'Counting Your Customers' the Easy Way: An Alternative to the Pareto/NBD Model," *Marketing Science*, 24 (2), 275-284.
- Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2006), "More Than Meets The Eye," *Marketing Research*, 18 (2), 8-15.
- Fader, Peter S., Bruce G. S. Hardie, and Jen Shang (2010), "Customer-Base Analysis in a Discrete-Time Noncontractual Setting," *Marketing Science*, 29 (6), 1086-1108.
- Gupta, Sunil (1988), "Impact of Sales Promotions on When, What, and How Much to Buy," *Journal of Marketing Research*, 25 (4), 342-355.
- Gupta, Sunil (1991), "Stochastic Models of Interpurchase Time with Time-Dependent Covariates," *Journal of Marketing Research*, 28 (1), 1-15.
- Gupta, Sunil, Dominique Hanssens, Bruce Hardie, William Kahn, V. Kumar, Nathaniel Lin, Nalini Ravishanker, and S. Sriram (2006), "Modeling Customer Lifetime Value," *Journal of Service Research*, 9 (2), 139-155.
- Hardie, Bruce G.S., Peter S. Fader, and Michael Wisniewski (1998), "An Empirical Comparison of New Product Trial Forecasting Models," *Journal of Forecasting*, 17 (3-4), 209-229.
- Hitt, Lorin M. and Frances X. Frei (2002), "Do Better Customers Utilize Electronic Distribution Channels? The Case of PC Banking," *Management Science*, 48 (6), 732-748.
- Hughes, Arthur (2005), *Strategic Database Marketing*, 3rd edn. New York: McGraw-Hill.
- Jain, Dipak C. and Naufel J. Vilcassim (1991), "Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach," *Marketing Science*, 10 (1), 1-23.
- Johnson, Norman L. and Samuel Kotz (1975), "On some generalized Farlie-Gumbel-Morgenstern distributions" *Communications in Statistics*, 4 (5), 415-427.
- Kumar, V. (2008), *Managing Customers for Profit: Strategies to Increase Profits and Build Loyalty*, 1st edn., Pearson Prentice Hall.

- Lewis, Michael (2004), "The Influence of Loyalty Programs and Short-Term Promotions on Customer Retention," *Journal of Marketing Research*, 41 (3), 281-292.
- Malthouse, Edward C. and Robert C. Blattberg (2005), "Can We Predict Customer Lifetime Value?" *Journal of Interactive Marketing*, 19 (1), 2-16.
- Moe, Wendy W. and Peter S. Fader (2004), "Capturing Evolving Visit Behavior in Clickstream Data," *Journal of Interactive Marketing*, 18 (1), 5-19.
- Montgomery, Alan L., Shibo Li, Kannan Srinivasan, and John C. Liechty (2004), "Predicting Product Purchase from Inferred Customer Similarity," *Marketing Science*, 23 (4), 579-595.
- Morrison, Donald G. and David C. Schmittlein (1980), "Jobs, Strikes, and Wars: Probability Models for Duration," *Organizational Behavior and Human Performance*, 25 (2), 224-251.
- Morrison, Donald G. and David C. Schmittlein (1988), "Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is It Worth the Effort?" *Journal of Business and Economic Statistics*, 6 (2), 145-159.
- Netzer, Oded, James Lattin, and V. Srinivasan (2008), "A Hidden Markov Model of Customer Relationship Dynamics," *Marketing Science*, 27 (2), 185-204.
- Park, Young-Hoon and Peter S. Fader (2004), "Modeling Browsing Behavior at Multiple Websites," *Marketing Science*, 23 (3), 280-303.
- Payne, Adrian and Pennie Frow (2005), "A Strategic Framework for Customer Relationship Management," *Journal of Marketing*, 69 (4), 167-176.
- Reinartz, Werner, Manfred Krafft, and Wayne D. Hoyer (2004), "The Customer Relationship Management Process: Its Measurement and Impact on Performance," *Journal of Marketing Research*, 41 (3), 293-305.
- Reinartz, Werner and V. Kumar (2000), "On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing," *Journal of Marketing*, 64 (4), 17-35.
- Sarmanov, O. V. (1966), "Generalized Normal Correlations and Two-Dimensional Frechet Classes," *Sovyet Mathematics*, 168, 596-599.
- Schmittlein, David C., Donald G. Morrison, and Richard Colombo (1987), "'Counting Your Customers': Who Are They and What Will They Do Next?" *Management Science*, 33 (1), 1-24.

- Schmittlein, David C. and Robert A. Peterson (1994), "Customer Base Analysis: An Industrial Purchase Process Application," *Marketing Science*, 13 (1), 41-67.
- Schweidel, David A. and Peter S. Fader (2009), "Dynamic Changepoints Revisited: An Evolving Process Model for New Product Sales," *International Journal of Research in Marketing*, 26 (2), 119-124.
- Schweidel, David A., Peter S. Fader, and Eric T. Bradlow (2008), "A Bivariate Timing Model of Customer Acquisition and Retention," *Marketing Science*, 27 (5), 829-843.
- Seetharaman, P. B. and Pradeep K. Chintagunta (2003), "The Proportional Hazard Model for Purchase Timing: A Comparison of Alternative Specifications," *Journal of Business and Economic Statistics*, 21 (3), 368-382.
- Singh, Siddharth S., Sharad Borle, and Dipak C. Jain (2009), "A Generalized Framework for Estimating Customer Lifetime Value When Customer Lifetimes are Not Observed," *Quantitative Marketing and Economics*, 7 (2), 181-205.
- Telang, Rahul, Peter Boatwright, and Tridas Mukhopadhyay (2004), "A Mixture Model for Internet Search-Engine Visits," *Journal of Marketing Research*, 41 (2), 206-214.
- Thomas, Jacquelyn S. (2001), "A Methodology for Linking Customer Acquisition to Customer Retention," *Journal of Marketing Research*, 38 (2), 262-268.
- Wedel, Michel, Wagner A. Kamakura, Wayne S. DeSarbo, and Frenkel Ter Hofstede (1995), "Implications for Asymmetry, Nonproportionality and Heterogeneity in Brand Switching from Piece-wise Exponential Mixture Hazard Models," *Journal of Marketing Research*, 32 (4), 457-462.
- Weinberg, Clarice Ring and Beth C. Gladen (1986), "The Beta-Geometric Distribution Applied to Comparative Fecundability Studies," *Biometrics*, 42 (3), 547-560.
- Wellan, Dee M. and A. S. C. Ehrenberg (1988), "A Successful New Brand: Shield," *Journal of the Market Research Society*, 30 (1), 35-44.
- Wheat, Rita D. and Donald G. Morrison (1990), "Assessing Purchase Timing Models: Whether or Not is Preferable to When," *Marketing Science*, 9 (2), 162-170.
- Yang, Sha and Greg M. Allenby (2003), "Modeling Interdependent Consumer Preferences," *Journal of Marketing Research*, 40 (3), 282-294.

CHAPTER 2

**VALUING CUSTOMERS CATEGORY BY CATEGORY: AN INTEGRATED
MODEL FOR MULTI-CATEGORY CUSTOMER LIFETIME VALUE
ANALYSIS**

2.1 Introduction

Customer lifetime value (CLV) is a key metric to manage and develop relationships with customers. As a disaggregate metric, CLV can be used to tailor marketing decisions to increase profitability at the customer level. For example, marketers often customize promotional offers based on the estimates of CLV and expend greater effort to retain customers with a high CLV (Kumar 2008). Beyond its use for marketing and resource allocation, when aggregated, CLV also serves as a good proxy for the overall value of a firm, supplementing financial measures of past performance (Gupta, Lehmann, and Stuart 2004).

Given the importance of measuring CLV, marketing researchers have developed a variety of models to predict the metric in different contexts. (See Gupta et al. (2006) and Fader and Hardie (2009) for relevant discussion.) These contexts are generally classified into contractual and non-contractual settings. Between the two cases, measuring CLV is known to be more challenging in a non-contractual setting because the lifetime of a customer is not observed (Bell et al. 2002, Fader, Hardie, and Lee 2005a, Singh, Borle, and Jain 2009). The “buy ‘til you die” framework is a widely adopted modeling approach for customer base and lifetime value analysis in a non-contractual setting. This class of models is based on the assumption that customers repeatedly conduct transactions at a firm while they are in an active state (alive), and at some unobserved time be-

come permanently inactive (dead). Though many variants of “buy ‘til you die” models have been proposed (e.g., Schmittlein, Morrison, and Colombo 1987, Fader, Hardie, and Lee 2005b, Abe 2009), they primarily focus on analyzing transactional patterns associated with a single product category or a firm-level activity, such as the times at which purchases are made at a particular retailer.

Oftentimes, however, customers’ shopping behavior involves the purchase of multiple categories at a single visit. In such cases, marketers can gain a more comprehensive understanding of their customer base by analyzing customers’ purchase patterns within individual product categories, and consequently assessing each category’s contribution to CLV. Such category-level analysis would be of particular importance to firms making marketing and operational decisions at the category level. Marketers could also benefit by examining customer shopping patterns across categories. In particular, the prediction of shopping basket choice and its relationship with CLV can be used in developing cross-selling strategies and making resource allocation decisions across categories within individual customers. Our objective of this research is therefore to propose an integrated modeling framework for multi-category CLV analysis in a non-contractual setting.

Modeling customer purchase patterns across categories requires a number of careful considerations. First, a customer’s arrival process to the firm and her category purchase decisions are likely to be interdependent. Imagine a customer whose shopping trips to the store are driven by two categories. The product categories in which she purchases on a given transaction may affect when she would visit the firm next. For example, if the two categories are complementary to each other for joint consumption, purchasing both categories together may re-

sult in a shorter time until the customer's next visit to the firm. Alternatively, if the categories are substitutes in which the consumption of one category delays the consumption of the other, the customer's next interarrival time after purchasing both categories may be longer compared to the case in which she had only purchased one of the product categories. The association of the interarrival time with category purchases may also arise from stockpiling, with large purchase amounts in one time period delaying the time of her next shopping trip (e.g., Jen, Chou, and Allenby 2009). In turn, the customer's interarrival time to the store may affect her category purchase decisions at the visit. Second, in a multi-category context, a customer's purchase decisions on a given shopping trip may be correlated across categories. Accounting for such associations is important to capture the simultaneous choice of categories in a shopping basket. Third, repeatedly observed shopping behavior from the same customer could be correlated over time, as a customer's current choice decisions are likely to be associated with past outcomes both within and across categories. Finally, latent customer defection needs to be considered in a non-contractual setting.

In this research, we develop a model for multi-category CLV analysis which accommodates the aforementioned complications by generalizing the "buy 'til you die" framework to a multi-category context. Toward this end, we jointly model customers' arrival process to the firm, multi-category purchase incidence and amount decisions, and latent defection while explicitly accounting for the shopping dynamics arising from the interplay of purchase timing and incidence across categories. The key challenge in modeling the timing process of customer arrivals and its association with multi-category purchase behavior is that the number of possible compositions of shopping baskets increases exponentially as the number of categories increases, and so does the complexity of the

model. To avoid this “curse of dimensionality,” we propose a novel modeling approach using a latent space of product categories. Our approach extends to multi-category cases with ease and yields a parsimonious timing model. We then model a customer’s decisions on category choice and purchase amount conditional on her visit to the firm. Our multi-category model accounts for the interdependent choices of categories and the correlation of repeatedly observed outcomes both within and across categories. We also allow customers’ category purchase decisions to depend on the timing of their past purchase of the category. Therefore, our model captures the sequential interaction between shopping baskets and interarrival times.

We apply our model to category-level customer transaction data from a leading beauty care company in Korea. We empirically assess the need to include the different modeling components and find support for the inclusion of each. As a result, the proposed model offers excellent fit and performance in predicting customers’ purchase patterns across categories, in addition to detailed inferences about their shopping behavior with respect to the timing, choice, and amount of category purchase. The results are then used to measure CLV at the category level and assess the contribution of each category to a firm-level CLV. The ability of our model to predict customers’ multi-category purchase patterns also allows us to examine the relationship between their shopping basket choice and CLV. We find that customers who purchase multiple categories together at the same transaction tend to have a higher CLV compared to those purchasing the categories separately over multiple visits. Our results can help marketers segment the customer base and yield insights for category-level targeting and cross-selling strategies.

The remainder of this chapter is organized as follows. Section 2.2 gives an overview of prior literature related to our research. In section 2.3, we provide a detailed specification of our model. Section 2.4 describes the data used in our empirical analysis. In section 2.5, we discuss the model results and illustrate insights afforded by analyzing customer purchase patterns at the category level. Finally, section 2.6 concludes and suggests future research directions.

2.2 Previous Research

This work is related to two main streams of research – predicting customer purchase patterns and lifetime values in a non-contractual setting, and capturing multi-category choice behavior. We briefly review relevant literature on both streams of research and discuss the contribution of our work relative to them.

Predicting customers' purchase patterns from a timing perspective has been of great interest to marketing researchers. One of the widely used modeling frameworks for this purpose is the negative binomial distribution (NBD) model which characterizes customers' interarrival times by an exponential distribution and accounts for heterogeneity in the time-invariant mean arrival rates by a gamma distribution (e.g., Ehrenberg 1988, Morrison and Schmittlein 1988). Several researchers have extended the NBD model to take into account nonstationarity in the customer arrival process (e.g., Fader, Hardie, and Huang 2004), interdependence between two correlated processes (e.g., Park and Fader 2004), and the effect of time-varying explanatory variables (e.g., Gupta 1991, Fader, Hardie, and Huang 2004). Another important extension of the NBD-based timing models was to incorporate latent customer defections for customer base

analysis in a non-contractual setting. One of the first models to explicitly address this challenge is the Pareto/NBD model by Schmittlein, Morrison, and Colombo (1987). The Pareto/NBD model assumes two independent exponential timing processes for customer arrivals and latent defections, respectively. By coupling the timing processes, the model assesses a customer's lifetime in a probabilistic manner. The BG/NBD model by Fader, Hardie, and Lee (2005b) serves the same purpose as the Pareto/NBD but is easier to implement and still yields similar forecasting performance. The development of these "buy 'til you die" models has facilitated academic research on measuring CLV in a non-contractual context (e.g., Schmittlein and Peterson 1994, Reinartz and Kumar 2000, Fader, Hardie, and Lee 2005a, Singh, Borle, and Jain 2009).

Our research is also based upon a growing body of work on multi-category choice behavior. This stream of research begins with meta-analyses to compare and contrast choice behavior across categories (e.g., Fader and Lodish 1990, Narasimhan, Neslin, and Sen 1996). Recognizing that a customer's purchase decisions across categories are not independent, researchers have developed models to capture cross-category dependence in shopping basket selection and understand the effect of marketing mix variables on multi-category choice decisions. Common modeling approaches for this purpose include the multivariate probit model (e.g., Manchanda, Ansari, and Gupta 1999, Chib, Seetharaman, and Strijnev 2002) and the conditional logit model (e.g., Russell and Petersen 2000, Moon and Russell 2008). A number of modifications and alternative specifications have been also proposed to study various phenomena of interest in a multi-category context (e.g., Mehta 2007, Niraj, Padmanabhan, and Seetharaman 2008, Song and Chintagunta 2006).

Another approach that has been employed in extant research to capture correlated outcomes is latent space models (e.g., DeSarbo and Wu 2001, Bradlow et al. 2005). By locating brands/products within a latent space, their relationships can be inferred from observable data such as joint product ownership (Kamakura, Ramaswami, and Srivastava 1991, Li, Sun and Wilcox 2005) or a search engine's ability to find specific webpages (Bradlow and Schmittlein 2000). As we will discuss, the use of a latent space model allows us to avoid the curse of dimensionality with multiple product categories.

Given the intuitive appeal of understanding both purchase timing and choice decisions, several models have been developed to model interarrival times and category (or brand) choices together (e.g., Wagner and Taudes 1986, Gupta 1988, Chintagunta 1999, Kumar, Venkatesan, and Reinartz 2008). Yet, little research considers customer purchase patterns across categories for multi-category CLV analysis in a non-contractual setting. One of the key contributions of this research is therefore to bridge the gap in the literature between customer base analysis and multi-category choice behavior.

It is worth noting that our proposed model may not be the only way of modeling interdependent customer purchase patterns across categories. For example, prior research has modeled the associated timing processes of different categories or transactional activities by using multivariate distributions such as the Farlie-Gumbel-Morgenstern family of distributions (e.g., Chintagunta and Haldar 1998) and the Sarmanov family of distributions (e.g., Park and Fader 2004, Schweidel, Fader, and Bradlow 2008). However, they did not take into account customer defection and thus have limitations for CLV analysis. Furthermore, perhaps due to the accompanying complexities of using the multivariate dis-

tributions, they focused on the bivariate applications of the models. In comparison, our model, using a latent space approach, extends to multi-category cases in a parsimonious manner. This is an important benefit from a practical standpoint, as customers often make purchases across several categories in their transactions with a firm and thus the analysis of purchase behavior needs to scale to a large number of categories.

2.3 Model Development

We build a model for multi-category CLV analysis in a non-contractual setting. Our proposed model consists of the following four modeling components: (1) the arrival process of a customer to a firm, (2) her choice decisions across multiple categories conditional on a visit, (3) the amounts she spends in each purchased category conditional on the choice of the category, and (4) latent customer defection. We first present the specification of the modeling components and discuss how they interact with each other. We then describe our computational approach to estimating the model.

2.3.1 Timing Model

During the period $(0, T]$, where 0 corresponds to the beginning of the model calibration period and T is the censoring point that corresponds to the end of the data period, we observe customer i who makes J_i visits to a firm at times $t_{i1}, t_{i2}, \dots, t_{iJ_i}$. At each visit to the firm, the customer makes purchase decisions across K product categories. We define a binary variable C_{ijk} to indicate whether or

not customer i purchases category k at her j th visit. Then, customer i 's shopping basket at the j th visit can be represented by a vector of category choice outcomes, $\{C_{ij1}, C_{ij2}, \dots, C_{ijK}\}$. It is important to note that, in our offline shopping context, we have $\sum_{k=1}^K C_{ijk} \geq 1$ for any visit, because a customer's visit to the firm is observed only when she purchases at least one category.¹

We first model customer i 's arrival process to the firm by specifying the timing behavior for a set of $J_i - 1$ interarrival times, $t_{i2} - t_{i1}, t_{i3} - t_{i2}, \dots, t_{iJ_i} - t_{i,J_i-1}$, and the right-censored observation $T - t_{iJ_i}$. We assume that customer i 's interarrival time between her j th and $(j+1)$ th visits follows an exponential distribution with arrival rate λ_{ij} . Then, the density function for the interarrival times and the survival function for the right-censored observations are given by:

$$f(t_{i,j+1}|t_{ij}; \lambda_{ij}) = \lambda_{ij} \exp\{-\lambda_{ij}(t_{i,j+1} - t_{ij})\} \text{ and } S(T|t_{iJ_i}; \lambda_{iJ_i}) = \exp\{-\lambda_{iJ_i}(T - t_{iJ_i})\}. \quad (2.1)$$

The exponential assumption has been widely adopted in the marketing literature because of its parsimony and performance (e.g., Schmittlein, Morrison, and Colombo 1987, Moe and Fader 2004, Fader, Hardie, and Lee 2005b).

When a customer repeatedly shops for multiple product categories at the firm, her shopping basket choice at a given visit may influence the time until her next shopping trip, depending on the relationship between the purchased categories. Whereas joint purchasing of products consumed together may accelerate consumption and shorten the time until the next visit, simultaneously purchasing products that are substitutes in their consumption may slow the time until the next purchase. To take into account this probable association of the customer arrival process with multi-category purchase decisions, we allow

¹In this regard, the terms "visits", and "transactions" are used interchangeably throughout this chapter. This research focuses on modeling customer purchase patterns across categories at an offline firm. However, our model can be easily modified to an online context, where customer visits often involve no purchase (i.e., $\sum_{k=1}^K C_{ijk} = 0$).

the customer's arrival rate to vary dynamically depending on the combination of categories she purchased at the prior visit.

While it is appealing to consider the effect of the customer's shopping basket choice on the arrival rate, one of the major complications is that the number of possible compositions of shopping baskets increases exponentially as the number of categories increases, and so does the complexity of the model. Note that, for the customer's each visit, there are $2^K - 1$ possible combinations of categories in her shopping basket, ranging from $\{1, 0, \dots, 0\}$ where she purchases the first category only to $\{1, 1, \dots, 1\}$ where she purchases all the categories considered. For example, when $K = 10$, we need to consider 1,023 ($= 2^{10} - 1$) different compositions of shopping baskets and their impact on the arrival rate. To deal with this dimensionality issue, we propose a novel modeling approach using a latent space of product categories which allows us to capture the association of the arrival rate with shopping basket selection in a parsimonious manner.

At the heart of our approach is to place categories in a latent space so that the arrival rate after purchasing a specific combination of categories is jointly determined by the relative positions of the categories in the space. Specifically, we model the effect of shopping basket composition on the arrival rate using the Euclidean distances between the categories and the origin in the space. To formalize this idea, we specify λ_{ij} as:

$$\lambda_{ij} = \lambda_i \left\| \sum_{k=1}^K C_{ijk} P_k \right\|^{-1} \exp(X'_{ij} \alpha), \quad (2.2)$$

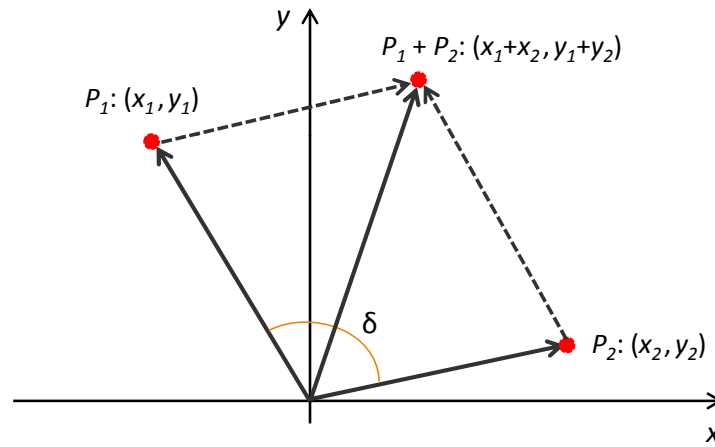
where λ_i is a customer-specific baseline arrival rate; P_k is the position of category k in an n -dimensional latent space and $\|\cdot\|$ denotes the Euclidean norm of a point (i.e., the Euclidean distance between the point and the origin)²; X_{ij} is a vector of

²We take an inverse of the Euclidian norm so that purchasing categories which are closer to

covariates which may affect λ_{ij} ; α is a vector of the corresponding parameters. In the empirical analysis of this research, X_{ij} consists of the lagged interarrival time (i.e., the customer's interarrival time between her $(j - 1)$ th and j th visits) and the customer's total purchase amount aggregated across all categories at her j th visit to consider their effects on the arrival rate.

We illustrate the intuition behind our use of the Euclidean norm in equation (2.2) with an example of two categories on a two-dimensional space. P_1 and P_2 in Figure 2.1 represent categories 1 and 2, respectively, on the space where the positions of the categories are denoted by their x - and y -coordinates. The evaluation of the Euclidean norm in equation (2.2) and the inverse relationship between the arrival rate and interarrival time indicate that the effect of purchasing only category k on the expected interarrival time for the next visit is given by $\|P_k\| = (x_k^2 + y_k^2)^{1/2}$. Thus, a larger (smaller) value of $\|P_k\|$ implies that the customer's interarrival time after purchasing category k tends to be longer (shorter).

Figure 2.1: Two-Category Representation in a Two-Dimensional Space



each other in the space (meaning they are more substitutable) likely leads to a longer interarrival time to the next visit, given the inverse relationship between the arrival rate and the mean interarrival time.

Next, when a customer purchases both categories 1 and 2, its effect on the expected interarrival time is given by the Euclidean norm of the summed vector, $\|P_1 + P_2\| = \{(x_1 + x_2)^2 + (y_1 + y_2)^2\}^{1/2}$. Note that the Euclidean norm varies depending on the angle, δ , between the two vectors, P_1 and P_2 . As δ decreases (increases), $\|P_1 + P_2\|$ increases (decreases) and thus the interarrival time after purchasing both categories becomes larger (smaller). Thus, δ reflects the relationship between the categories and their combined effect on the interarrival time (or the arrival rate). We expect that two substitutable categories would be close to each other on the latent space (i.e., small δ), resulting in a longer interarrival time after purchasing both categories compared to the case of purchasing one of them only. In contrast, if the categories are complementary with synergies from their joint consumption, they would be distant from each other on the space (i.e., large δ) as the co-purchasing of the categories would result in a shorter interarrival time. The same logic would apply to cases with more than two categories.

It is worth noting that when an one-dimensional space is employed, each category is represented by one coordinate on the space (i.e., x -coordinate only). Hence, equation (2.2) reduces to the specification which only accounts for the main effects of the categories. Alternatively, the latent space can be extended to more than two dimensions. Selection of the number of dimensions in the latent space would depend on the goal of the analysis. A two- or three-dimensional space would allow for a simple presentation of categories in the space. However, a higher dimensional space could allow more flexibility and thus a better fit, as the number of categories increases. Importantly, we need to restrict the maximum number of dimensions in the latent space to the number of categories considered, because the positions of the categories cannot be identified in a di-

dimensional space higher than the number of categories.

The Euclidean norms of a set of vectors are invariant under rotation and reflection of the space. Hence, there is an infinite number of positions of categories giving the same likelihood for the timing model. We therefore discuss identification conditions of the model. For a two-dimensional space, we first restrict the x -coordinate of category 1 to be positive and the y -coordinate of the category to zero. This takes into account rotation of the space. Second, we restrict the y -coordinate of category 2 to be positive. This takes into account reflection of the space over the y -axis. Finally, because λ_{ij} is proportional to the multiplication of λ_i and $\left\| \sum_{k=1}^K C_{ijk} P_k \right\|^{-1}$, there are infinite number of combinations of λ_i and P_k 's that give the same value of λ_{ij} . To consider this, we set the x -coordinate of category 1 to one. In summary, for our model with a two-dimensional space, we need $x_1 = 1$ and $y_1 = 0$ for category 1 and $y_2 > 0$ for category 2. No further identification condition is necessary on the coordinates of other categories when more than two categories are considered.

Similar identification conditions can be derived for a general case with an n -dimensional space as follows. First, we restrict the first coordinate of category 1 to one and all other coordinates of the categories to zero. For category k such that $1 < k \leq n$, we restrict the k th coordinate of the category to be positive and the m th coordinate to be zero for all m such that $m > k$. No identification condition is required on the coordinate of category $k > n$.

Finally, we discuss the benefit of our latent space approach that can parsimoniously capture the effect of shopping basket choice on the arrival rate. In Table 2.1, we compare the number of parameters required to consider the effect under our approach using the maximum-dimensional space (i.e., when the

number of dimensions in the space equals the number of categories considered) with the naive approach of specifying the effect using dummy variables (i.e., assigning a dummy variable for each composition of shopping baskets except the baseline case). For example, when five categories are considered, our approach requires maximum 14 parameters while the naive approach requires 31 dummy variables. Overall, the table shows that our approach allows us to approximate the effect of shopping basket choice with considerably fewer parameters.

Table 2.1: Number of Parameters Required to Consider the Effect of Shopping Basket Choice on Arrival Rate

No. of categories	Proposed approach (with the max.-dimensional space)	Naive approach
2	3	3
3	6	7
4	10	15
5	15	31
\vdots	\vdots	\vdots
10	55	1,023

2.3.2 Multi-category Choice Model

We next model a customer's multi-category choice behavior upon her visit to the firm. Our proposed choice model is a modified version of the conditional logit model. The conditional logit model has been employed in several marketing studies to account for the simultaneous and interdependent choices of multiple product categories (e.g., Russell and Petersen 2000, Moon and Russell 2008). In particular, the model allows us to explicitly capture cross-category dependence

by incorporating the choice outcomes in other categories into the choice decision in the focal category.

We specify the probability that customer i purchases category k at her j th visit conditional on her choices of other categories, using the following logit form:

$$\Pr(C_{ijk} = 1 | C_{ijk'}, k' \neq k) = \left[1 + \exp \left\{ - \left(\pi_{ijk} + \sum_{k' \neq k} \theta_{kk'} C_{ijk'} + \sum_{k'} \psi_{kk'} C_{i,j-1,k'} \right) \right\} \right]^{-1}, \quad (2.3)$$

where π_{ijk} captures the time-varying category-specific effect for category k at customer i 's j th visit to the firm and the term $\sum_{k' \neq k} \theta_{kk'} C_{ijk'}$ reflects the effects of the choices of other categories on the choice decision in category k . Hence, $\theta_{kk'}$ measures the degree of interdependence between category k and k' . A positive $\theta_{kk'}$ implies that the purchase of category k' tends to increase the probability of purchasing category k at the same transaction. A negative $\theta_{kk'}$ reflects a negative association between the choices of categories k and k' . Finally, the term $\sum_{k'} \psi_{kk'} C_{i,j-1,k'}$ captures the effects of the choices of categories in the prior transaction on the choice of category k in the current transaction. A positive (negative) $\psi_{kk'}$ implies that the purchase of category k' in the prior transaction increases (decreases) the probability of purchasing category k in the current transaction.

The probability of purchasing a category may also be influenced by the time of her last purchase of the category. To consider such an effect, we model the time-varying category-specific intercept π_{ijk} as:

$$\pi_{ijk} = \beta_{0ik} + \beta_{1k} \text{ET}_{ijk} + \beta_{2k} (\text{ET}_{ijk})^2, \quad (2.4)$$

where β_{0ik} is a customer-specific baseline purchase tendency for category k . Being constructed from customer purchase patterns in the data, the covariate ET_{ijk}

is the time between customer i 's j th visit and her last purchase of category k before the visit. Including ET_{ijk} and its quadratic term allow us to capture the possible nonlinear time effect of the prior category purchase on the current purchase decision. Thus, our model accounts for the sequential interdependence between shopping baskets and interarrival times.

Because the conditional probabilities in Equation (2.3) are mutually dependent across categories, we cannot use the equation to predict a customer's choice decisions across multiple categories. However, by assuming $\theta_{kk'} = \theta_{k'k}$ and applying the theorem by Besag (1974), it can be shown that the unconditional joint distribution of $\mathbf{C}_{ij} = \{C_{ij1}, C_{ij2}, \dots, C_{ijK}\}'$ is given by:

$$P\left(\mathbf{C}_{ij} = \mathbf{C}_{ij}^*, \sum_{k=1}^K C_{ijk}^* \geq 1\right) = \frac{\exp(\boldsymbol{\pi}_{ij}' \mathbf{C}_{ij}^* + \frac{1}{2} \mathbf{C}_{ij}^{*'} \boldsymbol{\Theta} \mathbf{C}_{ij}^* + \mathbf{C}_{i,j-1}' \boldsymbol{\Psi} \mathbf{C}_{i,j-1})}{\sum_{\mathbf{C}_{ij} \neq \{0,0,\dots,0\}} \exp(\boldsymbol{\pi}_{ij}' \mathbf{C}_{ij} + \frac{1}{2} \mathbf{C}_{ij}' \boldsymbol{\Theta} \mathbf{C}_{ij} + \mathbf{C}_{i,j-1}' \boldsymbol{\Psi} \mathbf{C}_{i,j-1})}, \quad (2.5)$$

where

$$\boldsymbol{\pi}_{ij} = \begin{bmatrix} \pi_{ij1} \\ \pi_{ij2} \\ \vdots \\ \pi_{ijK} \end{bmatrix}, \boldsymbol{\Theta} = \begin{bmatrix} 0 & \theta_{12} & \cdots & \theta_{1K} \\ \theta_{12} & 0 & \cdots & \theta_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{1K} & \theta_{2K} & \cdots & 0 \end{bmatrix} \text{ and } \boldsymbol{\Psi} = \begin{bmatrix} \psi_{11} & \psi_{12} & \cdots & \psi_{1K} \\ \psi_{21} & \psi_{22} & \cdots & \psi_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{K1} & \psi_{K2} & \cdots & \psi_{KK} \end{bmatrix} \quad (2.6)$$

and the summation in the denominator is across the possible shopping baskets constructed from K categories excluding the “no-purchase” case. The derivation of equation (2.5) is provided in the Appendix.

2.3.3 Purchase Amount Model

We model the amount spent by a customer within each category at a given visit. A customer's decision on purchase amount in a category is fully conditional

on her choice of the category. If the customer does not choose category k in her multi-category choice decision (i.e., $C_{ijk} = 0$), the purchase amount of the category is zero. To incorporate the conditionality and non-negativity of the amount, we specify customer i 's purchase amount of category k at her j th visit as follows:

$$\text{AMNT}_{ijk} = \begin{cases} \text{Log-normal}(m_{ijk}, \sigma_k^2) & \text{if } C_{ijk} = 1, \\ 0 & \text{otherwise.} \end{cases} \quad (2.7)$$

We further specify the time-varying mean parameter m_{ijk} to consider the effect of lagged purchase amount and the time elapsed since the last purchase of the category:

$$m_{ijk} = \gamma_{0ik} + \gamma_{1k} \text{LAMNT}_{ijk} + \gamma_{2k} \text{ET}_{ijk} + \gamma_{3k} (\text{ET}_{ijk})^2, \quad (2.8)$$

where γ_{0ik} is a customer-specific intercept for category k and LAMNT_{ijk} is the purchase amount in category k in the last purchase occasion for the category. Note that by definition, LAMNT_{ijk} is different from $\text{AMNT}_{i,j-1,k}$. LAMNT_{ijk} equals to $\text{AMNT}_{i,j-1,k}$ if the customer purchased category k at her $(j-1)$ th visit. Otherwise, LAMNT_{ijk} is not equal to $\text{AMNT}_{i,j-1,k}$ because $\text{AMNT}_{i,j-1,k} = 0$ and LAMNT_{ijk} is the amount spent in a prior purchase of category k before the $(j-1)$ th visit. Similarly with the multi-category choice model, the covariate ET_{ijk} and its quadratic term are included to consider the nonlinear time effect of the prior category purchase.

2.3.4 Customer Defection

The literature suggests the importance of taking into account latent customer defections for customer base and lifetime value analysis in a non-contractual

context (e.g., Schmittlein, Morrison, and Colombo 1987, Fader, Hardie, and Lee 2005a, Singh, Borle, and Jain 2009). In particular, failing to consider latent defection can result in the underestimation of customers' arrival rate in an active state and the overestimation of CLV. To account for latent customer defections, we assume that after any visit to the firm, customer i becomes permanently inactive with probability r_i , consistent with the individual-level attrition process of the BG/NBD model (Fader, Hardie, and Lee 2005b). The point at which the customer drops out is therefore distributed, according to a (shifted) geometric distribution with probability density:³

$$P(\text{customer } i \text{ drops out after her } j\text{th visit}) = r_i(1 - r_i)^{j-1}. \quad (2.9)$$

It is important to note that, in our multi-category context, purchasing cessation may occur at two distinct levels. First, attrition may occur at the firm level, as modeled in equation (2.9). This temporal dynamic affects purchasing behavior across all product categories. Another way in which purchasing may cease is at the category level. Should a customer decide to cease purchasing in a particular product category but continue to purchase in other product categories, this would manifest through the multi-category choice model in equation (2.3) with the covariate ET_{ijk} and its quadratic term included in equation (2.4). For

³The dropout probability may also vary across visits depending on the combination of categories the customer purchases at each visit. To consider this, we have applied the latent space approach and modeled the probability that customer i becomes permanently inactive after the j th transaction at the firm as:

$$r_{ij} = \left[1 + \exp \left(-\varphi_i \left\| \sum_{k=1}^K C_{ijk} Q_k \right\| \right) \right]^{-1},$$

where φ_i represents a customer-specific baseline dropout tendency and Q_k is a point representing category k on the latent space. Similar to the timing model of customer arrivals, the probability of customer defection is determined based on the relative positions of the purchased categories in the space. However, we have found that many estimates of the coordinates are insignificant and there is no significant improvement in model fit.

example, the scenario that a customer who made repeat purchases of a category stops buying the category at some point can be captured with the concave effect of ET_{ijk} on the probability of purchasing the category.

2.3.5 Customer Heterogeneity and Correlation Structure

In modeling a sequence of observations for each customer, we expect customers are heterogeneous in their shopping frequency and tendency of defection as well as multi-category purchase behavior. To incorporate customer heterogeneity into our model, we specify the customer-specific model parameters as follows. For the customer-specific baseline arrival rate in equation (2.2), we assume that λ_i follows a lognormal distribution to ensure that λ_i is positive. Second, we reparameterize the dropout probability r_i in equation (2.9) as $r_i = \frac{\exp(\omega_i)}{1+\exp(\omega_i)}$ to ensure $r_i \in [0, 1]$, and assume that ω_i follows a normal distribution. Third, we assume that the customer-specific intercepts $\beta_{0i1}, \beta_{0i2}, \dots, \beta_{0iK}$ and $\gamma_{0i1}, \gamma_{0i2}, \dots, \gamma_{0iK}$ in equations (2.4) and (2.8) follow a normal distribution. Taken together, we assume

$$\begin{bmatrix} \log \lambda_i \\ \omega_i \\ \beta_{0i1} \\ \vdots \\ \beta_{0iK} \\ \gamma_{0i1} \\ \vdots \\ \gamma_{0iK} \end{bmatrix} \sim \text{MVN} \left(\begin{bmatrix} \mu_\lambda \\ \mu_\omega \\ \mu_{\beta 1} \\ \vdots \\ \mu_{\beta K} \\ \mu_{\gamma 1} \\ \vdots \\ \mu_{\gamma K} \end{bmatrix}, \Sigma \right) \quad (2.10)$$

to allow for interdependence among the model parameters. Incorporating the covariance structure allows for dependencies across the outcome measures and makes an efficient use of information in the data.

2.3.6 Computational Approach

We adopt a Bayesian approach and use the Markov chain Monte Carlo (MCMC) methods to estimate the proposed model. Because the posterior distributions of some parameters are not standard, the model can be estimated using the Gibbs sampler (Gelfand and Smith 1990) with the Metropolis-Hastings steps (Hastings 1970). The samples obtained from the MCMC algorithm are then used to compute summary measures of the parameter estimates. Implementation of this method is relatively easy in the publicly available software WinBUGS. It is worth noting that the model can be fit in WinBUGS as it helps facilitate an uncomplicated implementation of our model and its use by practitioners who utilize the software. Further details of the estimation procedure are available from the authors on request.

Results reported are summarized from the output of three independent MCMC chains run for 40,000 iterations, each started from hyper-dispersed starting values, with a burn-in period of 20,000 iterations and utilizing the 60,000 draws (20,000 per chain) thereafter. To complete the Bayesian specification of the model, we assign priors to the model parameters. Because we lack any prior information, we take the usual route and assign noninformative conjugate priors to the parameters. For aggregate-level parameters and mean parameters, we use a normal density prior. For the variance-covariance matrix, we assume that the inverse of the matrix follows a Wishart distribution.

2.4 Data

The data used in this research come from one of the largest beauty care company in Korea.⁴ Our dataset consists of the transaction histories of 2,870 customers who purchased various beauty care products of a high-end brand designed for women at department stores in Korea from January 2008 to December 2010. The data contain information about the customers' shopping basket choice in each transaction, including the day of the transaction, products purchased, and price paid.

For our purpose, we categorized 79 cosmetics products (i.e., SKU) purchased by the customers during the data period into the following four categories: namely, (1) basics, (2) creams, (3) serums, and (4) makeups. This classification is also used on the firm's dashboard for marketing planning and performance evaluation. Basics consist of lotions and toners which are frequently purchased together as must-have skin care items for women. On average, products in this category are sold at a significantly lower price than other skin care products. Creams and serums are advanced and topical (e.g., anti-aging and anti-wrinkle) skin care products. Finally, Makeups contain products for facial makeup and cleansing.

Table 2.2 summarizes the descriptive statistics of our data. Depending on the time of the first transaction at the firm, the length of the observation period ranges from 188 to 1,094 days across customers with the mean of 724 days. On average, customers made 9.1 transactions with the firm and purchased 1.6 different categories per transaction. Out of the total transactions by the customers,

⁴Given our confidentiality agreement with the company, we are unable to provide more details about the company.

54% were the purchase of only one category, 30% involved two categories, 13% involved three categories, and the remaining 3% were the purchase of all four categories. Among the categories, creams were most frequently purchased (4.6 times per customer) and makeups were least purchased (2.4 times per customer) over the data period. In terms of purchase amount per transaction, customers spent most in creams (\$98.7) and least in makeups (\$15.7).⁵

Table 2.2: Descriptive Statistics of the Data

	Mean	Std. Dev.
<u>Across customers</u>		
Observation duration (days)	723.83	314.89
No. of transactions	9.06	5.19
No. of purchases of basics	3.84	2.97
No. of purchases of creams	4.62	3.01
No. of purchases of serums	3.97	3.03
No. of purchases of makeups	2.46	2.71
<u>Across transactions</u>		
No. of categories purchased	1.64	0.80
Purchase amount of basics	51.73	185.27
Purchase amount of creams	98.71	146.60
Purchase amount of serums	61.63	109.85
Purchase amount of makeups	15.78	39.87

Given the customers' purchases of multiple categories per transaction, we compute how many times each shopping basket was purchased, the number of customers who purchased the shopping basket, and the average time taken until the next visit since they purchased the shopping basket in Table 2.3. The first column of the table lists all possible compositions of shopping baskets, de-

⁵All transactions were recorded in Korean currency (won). At the time our data were collected, \$1 corresponded approximately to 1,200 Korean won.

noted by a vector whose four elements indicate the purchase of basics, creams, serums, and makeups in the order, respectively. For example, $\{1,0,0,0\}$ in the first row denotes a shopping basket containing basics only. The table shows that the number of purchases varies considerably across shopping baskets from 480 for $\{1,1,0,1\}$ to 4,489 for $\{0,1,0,0\}$. We find that all combinations of categories were purchased by only a subset of the customers. The most frequently purchased shopping basket $\{0,1,0,0\}$ was purchased by 62% ($= 1,789/2,870$) of the customers and the least purchased shopping basket $\{1,0,1,1\}$ was purchased by only 12% ($= 339/2,870$) of them. Lastly, the average time taken until the next visit ranges from 66.3 days for $\{0,0,0,1\}$ to 95.4 days for $\{1,1,1,0\}$. Hence, the mean interarrival times differ up to one month depending on the combination of categories purchased, which clearly supports the need to consider a possible association of the arrival rate with prior shopping basket choice in modeling customer purchase patterns across categories.

Table 2.3: Shopping Basket and Interarrival Time

Shopping basket	No. of observations	No. of customers who purchased the shopping basket	Time taken until the next visit (days)
{1,0,0,0}	3,579	1,457	73.6
{0,1,0,0}	4,489	1,789	73.6
{0,0,1,0}	3,383	1,418	70.8
{0,0,0,1}	2,569	1,123	66.3
{1,1,0,0}	1,956	1,130	83.6
{1,0,1,0}	1,506	919	79.4
{1,0,0,1}	559	387	71.9
{0,1,1,0}	2,149	1,228	84.4
{0,1,0,1}	975	661	73.6
{0,0,1,1}	712	492	71.0
{1,1,1,0}	1,880	1,151	95.3
{1,1,0,1}	480	349	80.0
{1,0,1,1}	430	339	80.2
{0,1,1,1}	694	477	84.6
{1,1,1,1}	640	462	87.8

We use the first 30 months of data for model calibration and the remaining six-month period for model validation. To allow for shorter MCMC run times in our empirical analysis, without loss of generality, we created a systematic sample of our data. Specifically, we sorted the 2,870 customers by the number of their visits and the mean number of categories purchased per transaction as primary and secondary dimensions; then we selected every other panelist who made repeated transactions during the calibration period. This results in 1,311 customers in our sample data.

2.5 An Empirical Application

In this section, we provide an empirical demonstration of our model using our data. There are four major areas that we report upon in summarizing our results. First, we compare several benchmark models to show the superiority of the proposed model. Second, we examine the fit of our model with respect to its ability to predict customer purchase patterns across categories. Third, we describe inferences based on the posterior distributions of parameter estimates. Finally, we demonstrate the ability of our model to evaluate the contribution of individual categories to CLV and the relationship between the metric with multi-category purchase behavior.

2.5.1 Selecting the Number of Dimensions in the Latent Space

We first select the number of dimensions in the latent space for the timing model of customer arrival process. To provide a comparison of the timing models with

different dimensional latent spaces, we compute the log marginal density. In doing so, we exclude the multi-category choice model and the purchase amount model, because the modeling components are invariant with the number of dimensions in the latent space. As another model fit measure, we compute the mean absolute error (MAE) with respect to the number of visits to the firm, averaged across customers.

Table 2.4 shows the values of the log marginal density and MAE for the models with one-, two-, three-, and four-dimensional latent spaces, respectively. The results reveal that the model fit improves as the number of dimensions in the space increases from one to four, and thereby the timing model with a four-dimensional latent space provides the best fit with the highest log marginal density and the smallest MAE in both in-sample and out-of-sample periods. However, looking at the overall changes of the measures, we find that the improvement in model fit is only marginal between three- and four-dimensional spaces. We choose to use a three-dimensional space for the timing model because results are not qualitatively different between three- and four-dimensional spaces. Thus, the additional parsimony outweighs the benefit from the minor improvement in the model fit.

Table 2.4: Selection of the Number of Dimensions in the Latent Space

No. of dimensions	Log marginal density	In-sample MAE	Out-of-sample MAE
1	-30,282	3.13	1.34
2	-30,091	2.92	1.20
3	-29,961	2.77	1.09
4	-29,926	2.71	1.05

2.5.2 Model Comparison

Our proposed model is quite general and nests several models as special cases. We fit a number of nested models to investigate if the inclusion of different model components is warranted. The alternative models we examine vary with respect to two aspects: (1) whether or not the model incorporates the effect of shopping basket choice on the arrival rate and (2) whether or not the model accounts for the interdependence across category choices within a transaction.

The first benchmark model (Model 1) is the simplest model which assumes that a customer's arrival rate does not vary based on her shopping basket choice and choice decisions are independent across categories. This model is formulated by specifying the arrival rate λ_{ij} in equation (2.2) as $\lambda_{ij} = \lambda_i \exp(X'_{ij}\alpha)$, and assuming all $\theta_{kk'}$ in equation (2.3) are equal to zero. Note that, in this model, the choice probabilities are independent across categories and thus the firm-level timing process is thinned into four independent category-level timing processes. The second and third benchmark models (Models 2 and 3) generalize Model 1 by incorporating one of the modeling components. Model 2 accounts for the impact of shopping basket choice on the arrival rate but does not reflect the interdependence across category choices within a transaction, i.e., $\theta_{kk'} = 0$ in equation (2.3). Model 3 considers the cross-category effects in category choices but ignores the association of the arrival rate with shopping basket composition, i.e., $\lambda_{ij} = \lambda_i \exp(X'_{ij}\alpha)$ in equation (2.2). Finally, Model 4 is our full proposed model.

We compare these four benchmark models based on the log marginal density and MAE with respect to the number of purchases of each shopping basket by individual customers, averaged across shopping baskets and customers. Table

2.5 reports the model fit measures of the alternative models. It shows that Model 4, our proposed model, has the highest log marginal density and the smallest MAE in both in-sample and out-of-sample periods, and thus provides a better fit than other alternative models. The results suggest a probable association between the arrival rate and shopping basket choice, and interdependence across category choices within a transaction. Given the empirical evidence which supports the inclusion of the corresponding modeling components, we focus on Model 4 hereafter.

Table 2.5: Model Comparison

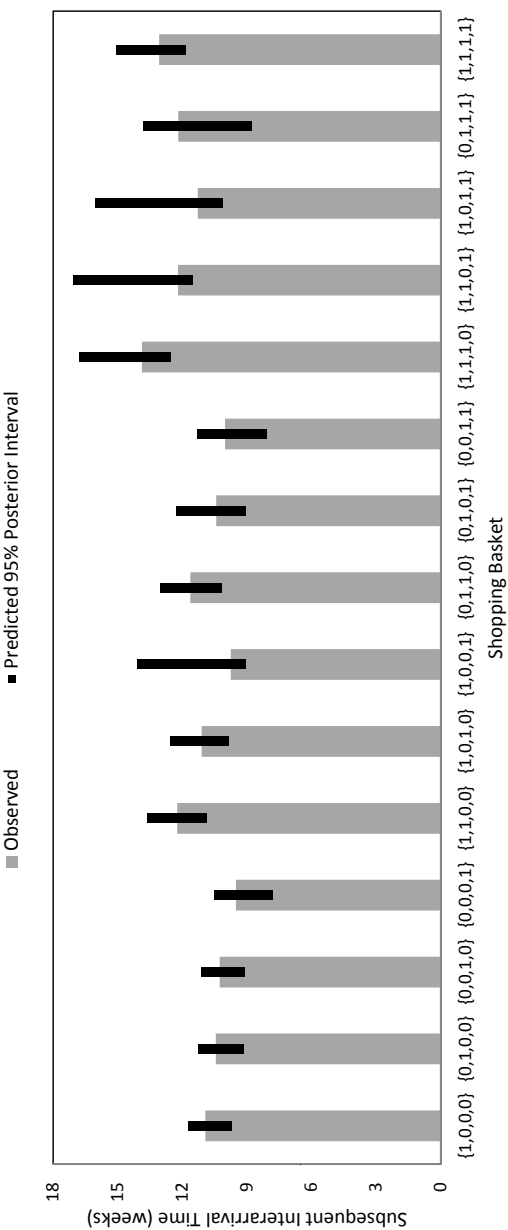
	Effect of shopping basket on the arrival rate	Interdependence across category choices	Log marginal density	In-sample MAE	Out-of-sample MAE
Model 1	×	×	-63,790	0.59	0.26
Model 2	✓	×	-63,337	0.57	0.24
Model 3	×	✓	-63,189	0.54	0.23
Model 4	✓	✓	-62,772	0.52	0.21

2.5.3 Model Fit

We assess whether the model properly captures the customers' purchase patterns across the categories. To this end, we predict the following four outcome measures, computed as posterior means across MCMC iterations: (1) the mean interarrival times after purchasing a shopping basket, (2) the cumulative number of purchases of a shopping basket, (3) the mean purchase amount of a category, and (4) the cumulative purchase amount of a category over the data period. As shown below, overall results indicate a good fit and predictive performance of our model.

In Figure 2.2, we compare the observed and predicted values of the mean interarrival times after the customers purchase each combination of categories. The results show that all of the observed values are contained in the 95% posterior intervals of their respective prediction values. The average of the percentage errors weighted by the number of observations for shopping baskets is 6.4%. While the overall model fit appears to be good across shopping baskets, the prediction error and the 95% posterior interval tend to be smaller for more frequently observed shopping baskets (e.g., $\{1,0,0,0\}$, $\{0,1,0,0\}$) and larger for less frequently observed ones (e.g., $\{1,0,0,1\}$, $\{1,1,0,1\}$). (See Table 2.3 for the number of observations for individual shopping baskets.)

Figure 2.2: Model Fit on Interarrival Time after Purchasing a Shopping Basket

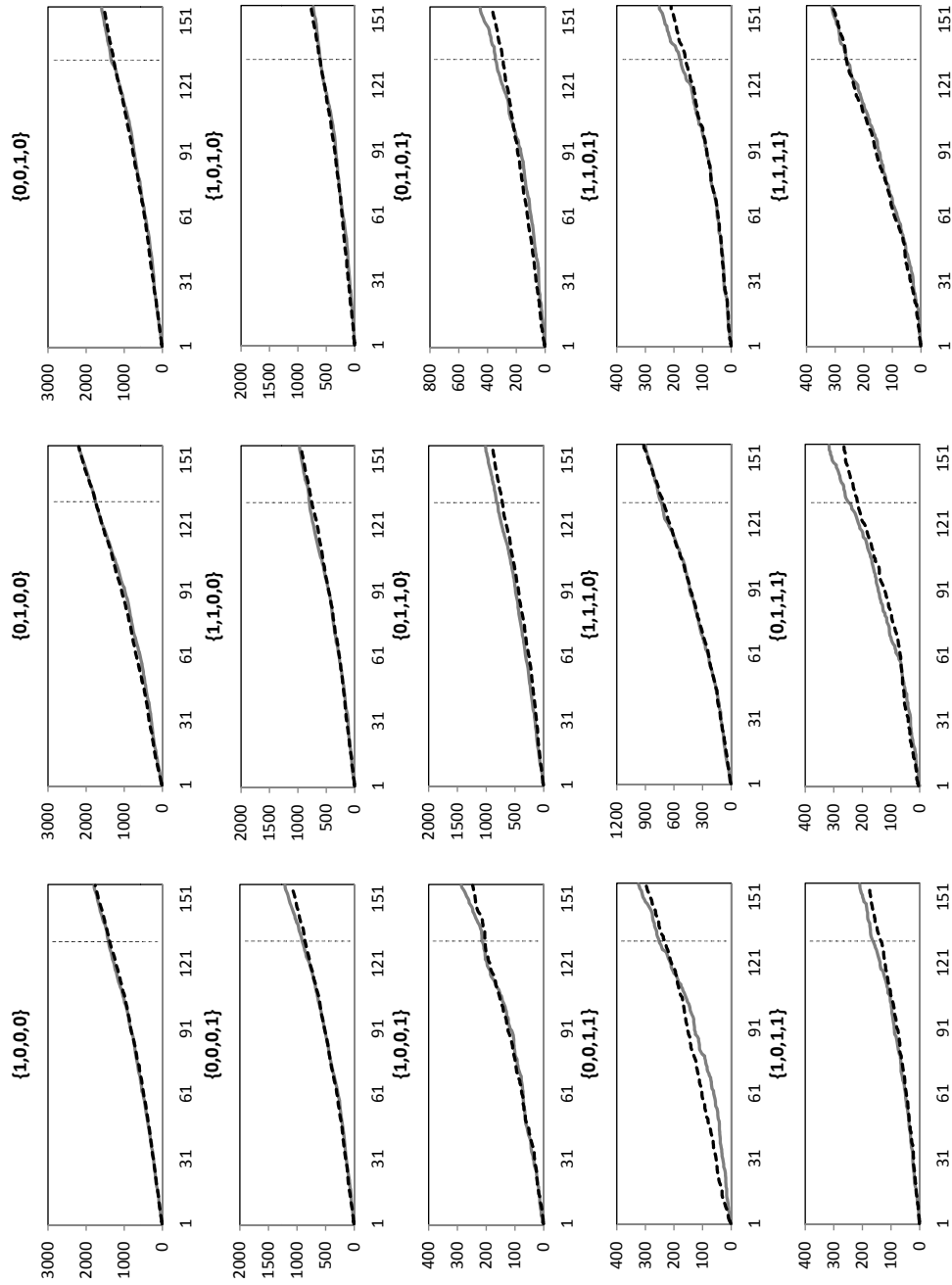


We next predict the cumulative number of purchases of each shopping basket over the data period. Figure 2.3 compares the mean values of the observed development of the measures and the predictions derived from our model. At the end of the calibration (validation) period, the model predicts the cumulative purchases at a 6.6% (8.2%) error rate, averaged across shopping baskets. We also find that for all shopping baskets, the observed values are contained in the 95% posterior intervals of their respective prediction values throughout the data period. As these tracking plots suggest, our model accurately tracks customers' purchases of shopping baskets as well as categories.

To examine the fit of the purchase amount model, we predict customers' mean purchase amount of a category conditional on their choice of the category. Figure 2.4 compares the distributions of the observed and predicted values. The fit is reasonable but there are a few peaks the model does not detect in predicting the purchase amount of a category, in particular, serums and makeups. In our data, those peaks are observed because customers' purchase amounts within the categories are concentrated on some discrete values linked to the retail prices of frequently purchased items. The fit could be improved by employing mixture distributions and adding additional parameters, but given the main purpose of this research, we choose to forgo this extension to maintain model parsimony.

Lastly, by considering all outcome measures of purchase time, choice, and amount across categories, we predict customers' cumulative purchase amount of a category over the data period. Figure 2.5 shows the comparison of the distributions of the observed and predicted values. The fit results clearly support the ability of our model to forecast the cumulative purchase amounts of individual categories and thus measure CLV at the category level.

Figure 2.3: Model Fit on Cumulative Purchases of a Shopping Basket



Note: (1) In all charts, the x-axis is week and the y-axis is the cumulative purchases of a shopping basket.
(2) The solid and dotted lines represent the observed and predicted values, respectively.

Figure 2.4: Model Fit on Mean Purchase Amount

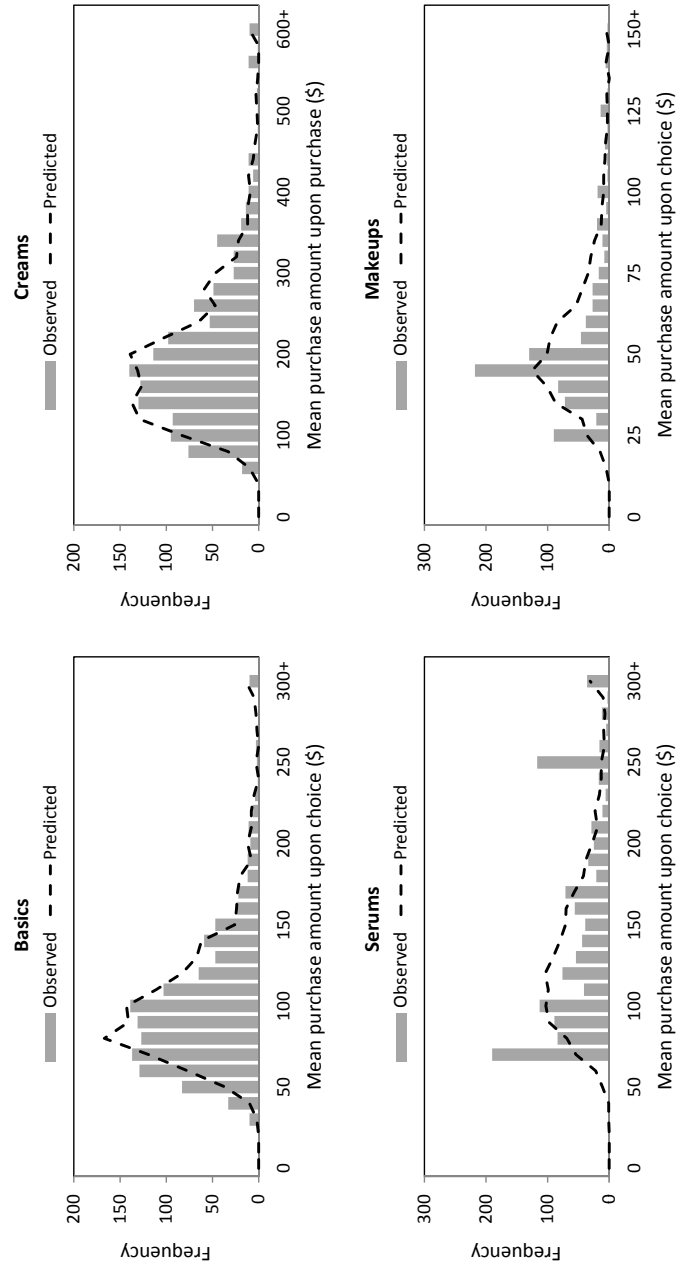
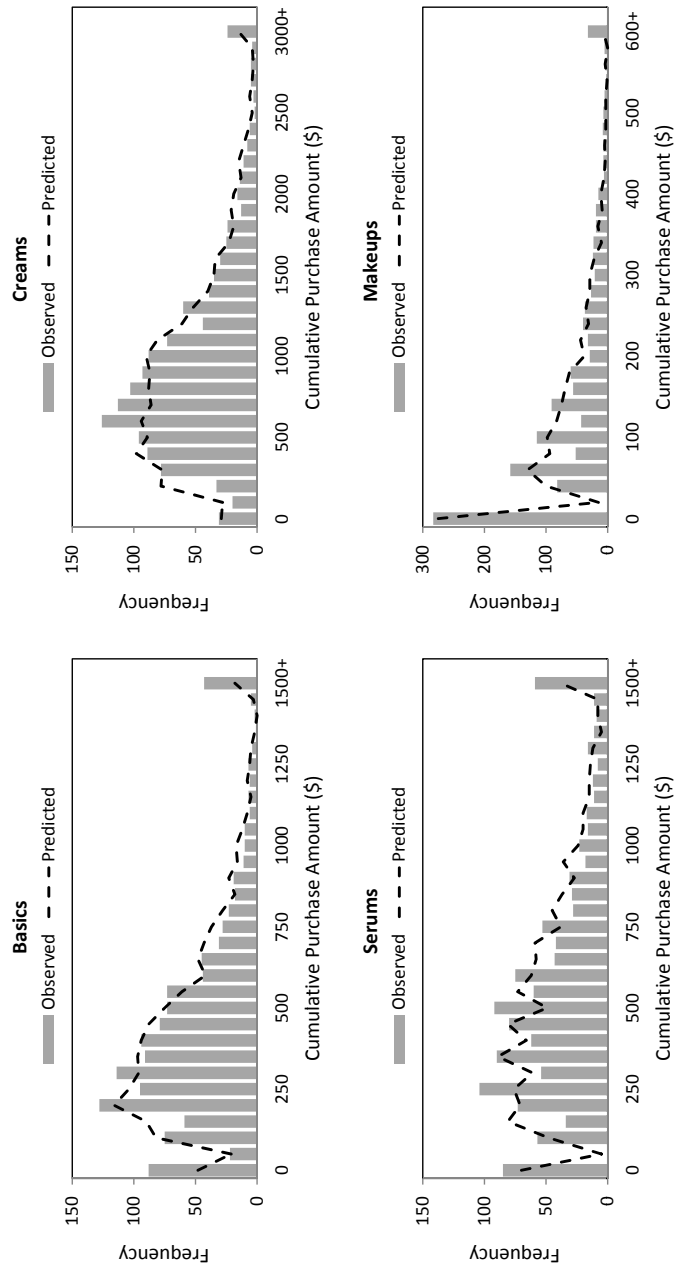


Figure 2.5: Model Fit on Cumulative Purchase Amount



2.5.4 Parameter Inferences

We describe inferences based on the estimates of our model parameters, obtained by looking at their posterior distributions. We begin with the results for the model of customer arrivals and defections. From the estimates of μ_λ , μ_ω , $\Sigma_{\lambda\lambda}$, and $\Sigma_{\omega\omega}$ in Table 2.6, we find that the mean baseline arrival rate is 0.09 ($= e^{\mu_\lambda + \Sigma_{\lambda\lambda}/2}$), and on average customers defect with a probability of 0.01 after their each visit to the firm.⁶ The estimate of α_1 indicates that the effect of the lagged interarrival time on the arrival rate is positively significant, implying that the interarrival times are negatively associated across visits. The estimate of α_2 suggests that a customer's total purchase amount at a given visit does not significantly affect her next arrival rate.

Table 2.6: Parameter Estimates of the Timing Model: Baseline Arrival Rate and Customer Defection

Parameter	Posterior mean	95% Posterior interval
μ_λ	-2.53	[-2.62,-2.45]
μ_ω	-4.82	[-5.09,-4.53]
$\Sigma_{\lambda\lambda}$	0.32	[0.11, 0.51]
$\Sigma_{\omega\omega}$	0.28	[0.14, 0.41]
α_1	3.70×10^{-4}	[2.21×10^{-4} , 5.08×10^{-4}]
α_2	-1.74×10^{-3}	[-3.86×10^{-3} , 4.36×10^{-4}]

Next, Table 2.7 reports the estimated coordinates of the four categories in the three-dimensional latent space employed in the timing model. As discussed, these estimates allow us to use information about customers' current shopping basket choice to predict the timing of their next transaction. The three axes com-

⁶Note that if $X \sim N(\mu, \sigma^2)$, $Y = e^X$ follows a log-normal distribution whose mean is given by $e^{\mu + \sigma^2/2}$.

prising the latent space could be interpreted based on the relative positions of the four categories in the space. For example, basics and makeups (creams and serums) are relatively close to each other with respect to their first coordinates. In this regard, the first axis may represent the characteristics of the categories which contrast basics and makeups to creams and serums in women's use of cosmetics, such as "essential versus optional" or "basic versus functional." Similarly, the second and third axes could be named "creams versus non-creams" and "serums versus non-serums," respectively. The second and third dimensions may thus capture differences in these categories not reflected in their position of the first coordinate.

Table 2.7: Parameter Estimates of the Timing Model: Coordinates of Categories in a Three-Dimensional Latent Space

	Coordinates		
	1st	2nd	3rd
Basics	1.00	0.00	0.00
	—	—	—
Creams	-0.33	1.03	0.00
	[-0.47,-0.20]	[0.93, 1.12]	—
Serums	-0.35	-0.26	0.95
	[-0.48,-0.21]	[-0.38,-0.14]	[0.86, 1.07]
Makeups	0.88	-0.10	-0.10
	[0.80, 0.97]	[-0.28, 0.09]	[-0.31, 0.10]

Note: The numbers in brackets denote the 95% posterior interval.

A second set of inferences arises from the multi-category choice model. Table 2.8 reports the estimates of parameters which capture the category-specific effects specified in equation (2.4). Based on the estimates of μ_k 's, we find that, among the four categories, customers' average baseline purchase tendency is highest for creams and lowest for makeups. Next, for all categories, the sig-

nificant estimates of β_{2k} 's and β_{3k} 's indicate a nonlinear effect of the elapsed time since the last purchase of a category (ET_{ijk}) on the probability of purchasing the category. In particular, we find that the overall effect of the elapsed time, $\beta_{2k}ET_{ijk} + \beta_{3k}(ET_{ijk})^2$, is positive up to some point (e.g., about 40 weeks for serums), implying that a longer period of non-purchase of a category leads to a higher probability of purchasing the category. However, when the elapsed time is beyond the time point, it has a negative effect on the purchase probability. As discussed earlier, this concave effect captures a category-level dynamic in the purchase tendency, enabling the model to account for category-level attrition as well as firm-level attrition.

Table 2.8: Parameter Estimates of the Choice Model: Category-Specific Effects

	Parameter	Posterior mean	95% Posterior interval
Basics	$\mu_{\beta 1}$	-1.46	[-1.60,-1.31]
	β_{11}	0.03	[0.02, 0.03]
	β_{21}	-2.40×10^{-4}	$[-3.20 \times 10^{-4}, -1.60 \times 10^{-4}]$
Creams	$\mu_{\beta 2}$	-0.98	[-1.12,-0.83]
	β_{12}	0.02	[0.01, 0.03]
	β_{22}	-1.98×10^{-4}	$[-3.01 \times 10^{-4}, -1.08 \times 10^{-4}]$
Serums	$\mu_{\beta 3}$	-1.59	[-1.76,-1.45]
	β_{13}	0.02	[0.02, 0.03]
	β_{23}	-2.88×10^{-4}	$[-3.80 \times 10^{-4}, -2.02 \times 10^{-4}]$
Makeups	$\mu_{\beta 4}$	-2.04	[-2.21,-1.82]
	β_{14}	0.01	[0.00, 0.02]
	β_{24}	-1.18×10^{-4}	$[-1.97 \times 10^{-4}, -4.12 \times 10^{-5}]$

Our next results are based on the estimates of $\theta_{kk'}$'s which capture the cross-category effects within a transaction. Table 2.9 shows that all parameter estimates, except the ones for the pairs of basics and makeups and of creams and

makeups, are positively significant, suggesting that the purchase of category k increases the probability of purchasing other categories that are positively tied with the category. We find that, among the four significantly associated pairs of categories, basics and serums (serums and makeups) have the largest (smallest) positive impact on each other. In particular, we find that the cross-category effects tend to be much weaker when makeups are involved (the third column of Table 2.9) than those between other categories (the first and second columns of Table 2.9), implying the purchase of makeups is less influenced by the choice of other categories. This might be because makeups are relatively more distinct from the other three categories which are all used for skin care purposes.

Table 2.9: Parameter Estimates of the Choice Model: Cross-Category Effects

	Creams	Serums	Makeups
Basics	0.58 [0.48, 0.67]	0.79 [0.70, 0.87]	0.10 [-0.01, 0.21]
Creams		0.76 [0.67, 0.86]	0.09 [-0.03, 0.20]
Serums			0.36 [0.23, 0.48]

Note: The numbers in brackets denote the 95% posterior interval.

The estimates of $\psi_{kk'}$'s for the lagged effects of category purchases are presented in Table 2.10. The resulting inferences could help marketers perform category-level targeting using customers' prior purchase records. For example, from the diagonal entries in the table, we find that for basics, creams, and serums, the purchase of a category in the prior transaction decreases the probability of purchasing the same category in the current transaction. In contrast to the three categories, the lagged purchase of makeups does not significantly

influence the likelihood of purchasing the category. In terms of lagged cross-category effects, the parameter estimates in the first (second) row of the table reveal that the lagged purchase of basics (creams) has a negatively significant effect on the purchase of creams (basics and serums). From the third row of the table, we find that the lagged purchase of serums does not affect the purchase of all other categories. The last row of the table indicates that the purchase of makeups decreases the probabilities of choosing basics and creams in the next transaction. Lastly, by looking at the table column-by-column, we find that for basics, creams, and serums, the purchase of a category is significantly affected by the purchase of the category as well as other categories in the prior transaction. In contrast, the choice of makeups is not influenced by the lagged purchase of any categories. Again, such distinction between makeups and the other three categories might be attributed to the characteristics of the products (i.e., makeups versus skin care items).

Table 2.10: Parameter Estimates of the Choice Model: Lagged Effects

	Basics	Creams	Serums	Makeups
Basics	-0.26 [-0.38,-0.13]	-0.08 [-0.20,-0.04]	-0.02 [-0.13, 0.10]	-0.04 [-0.17, 0.09]
Creams	-0.17 [-0.29,-0.05]	-0.32 [-0.43,-0.20]	-0.14 [-0.25,-0.03]	-0.14 [-0.28, 0.01]
Serums	-0.06 [-0.18, 0.06]	-0.01 [-0.14, 0.11]	-0.31 [-0.43,-0.19]	0.07 [-0.08, 0.23]
Makeups	-0.28 [-0.43,-0.14]	-0.32 [-0.45,-0.18]	-0.12 [-0.26, 0.03]	0.14 [-0.02, 0.30]

Note: The numbers in brackets denote the 95% posterior interval.

Another set of inferences is drawn from the purchase amount model. Table 2.11 reports the estimates of the parameters specified in equation (2.8). From the

estimates of γ_{1k} 's, we find that for basics and makeups, the purchase amount of a category has a positive dependence on the lagged purchase amount of the category (LAMNT_{ijk}). In contrast, the purchase amounts of creams are negatively correlated across transactions of the category. The estimates of γ_{2k} 's and γ_{3k} 's indicate that the purchase amount of creams varies significantly depending the elapsed time since the last purchase of the category (ET_{ijk}). Overall, however, we find that for all categories, both effects of lagged purchase amount and elapsed time on the purchase amount are very marginal by comparing the magnitudes of the covariates and their corresponding parameter estimates to the mean values of the category-specific intercepts, $\mu_{\gamma k}$'s.

Table 2.11: Parameter Estimates of the Purchase Amount Model

	Parameter	Posterior mean	95% Posterior interval
Basics	$\mu_{\gamma 1}$	4.48	[4.45, 4.52]
	γ_{11}	3.31×10^{-4}	[2.12×10^{-4} , 4.42×10^{-4}]
	γ_{21}	-1.23×10^{-3}	[-3.30×10^{-3} , 7.09×10^{-4}]
	γ_{31}	2.32×10^{-5}	[-2.99×10^{-6} , 5.05×10^{-5}]
Creams	$\mu_{\gamma 2}$	4.99	[4.95, 5.02]
	γ_{12}	-1.95×10^{-4}	[-3.45×10^{-4} , -4.66×10^{-5}]
	γ_{22}	7.17×10^{-3}	[4.87×10^{-3} , 9.57×10^{-3}]
	γ_{32}	-7.00×10^{-5}	[-1.04×10^{-4} , -3.71×10^{-5}]
Serums	$\mu_{\gamma 3}$	4.77	[4.74, 4.80]
	γ_{13}	5.87×10^{-5}	[-1.20×10^{-4} , 2.34×10^{-4}]
	γ_{23}	5.68×10^{-4}	[-1.56×10^{-3} , 2.39×10^{-3}]
	γ_{33}	-1.64×10^{-5}	[-4.39×10^{-5} , 1.17×10^{-5}]
Makeups	$\mu_{\gamma 4}$	3.86	[3.81, 3.90]
	γ_{14}	6.84×10^{-4}	[4.01×10^{-5} , 1.27×10^{-3}]
	γ_{24}	-3.46×10^{-4}	[-2.66×10^{-3} , 1.91×10^{-3}]
	γ_{34}	-2.98×10^{-6}	[-2.93×10^{-5} , 2.45×10^{-5}]

Finally, we report the estimated variance-covariance matrix Σ in Table 2.12.

The estimates of the variances of customer-specific parameters, in comparison to their mean values, suggest that customers are relatively more heterogeneous in category choice behavior, compared to their visit frequency and purchase amount of a category. The covariance estimates imply that customers' behaviors in purchase timing, choice, and amount tend to be significantly correlated both within and across categories. The results therefore demonstrate the validity of our proposed joint modeling framework.

Table 2.12: Estimated Σ

	$\log \lambda_i$	ω_i	β_{0i1}	β_{0i2}	β_{0i3}	β_{0i4}	γ_{0i1}	γ_{0i2}	γ_{0i3}	γ_{0i4}
$\log \lambda_i$	0.32	-0.12	-0.10	-0.14	-0.09	0.00	-0.06	-0.07	-0.07	-0.04
ω_i	[0.11, 0.51]	[-0.16, -0.08]	[-0.15, -0.07]	[-0.18, -0.10]	[-0.13, -0.05]	[-0.04, 0.05]	[-0.07, -0.04]	[-0.08, -0.05]	[-0.09, -0.06]	[-0.06, -0.02]
β_{0i1}		0.28	0.24	0.09	0.21	-0.07	0.10	0.09	0.09	0.06
		[0.14, 0.41]	[0.02, 0.43]	[0.00, 0.22]	[0.10, 0.38]	[-0.22, 0.15]	[0.06, 0.13]	[0.05, 0.13]	[0.05, 0.14]	[0.01, 0.09]
β_{0i2}			0.89	0.24	0.01	0.11	0.10	-0.02	0.00	0.01
			[0.72, 1.09]	[0.13, 0.36]	[-0.10, 0.13]	[-0.03, 0.26]	[0.06, 0.15]	[-0.06, 0.01]	[-0.03, 0.04]	[-0.03, 0.05]
β_{0i3}				0.83	0.18	0.44	-0.01	0.09	0.06	0.04
				[0.68, 1.03]	[0.06, 0.30]	[0.31, 0.60]	[-0.04, 0.03]	[0.05, 0.13]	[0.03, 0.09]	[0.01, 0.08]
β_{0i4}					0.94	0.18	-0.03	0.06	0.02	0.03
					[0.74, 1.16]	[0.01, 0.35]	[-0.07, 0.01]	[0.02, 0.10]	[-0.02, 0.05]	[-0.01, 0.07]
γ_{0i1}						0.80	-0.05	-0.01	0.02	0.00
						[0.63, 1.06]	[-0.09, -0.01]	[-0.05, 0.02]	[-0.01, 0.06]	[-0.04, 0.05]
γ_{0i2}							0.12	0.05	0.05	0.04
							[0.11, 0.14]	[0.04, 0.06]	[0.04, 0.07]	[0.02, 0.05]
γ_{0i3}								0.12	0.09	0.03
								[0.11, 0.14]	[0.07, 0.10]	[0.02, 0.04]
γ_{0i4}									0.11	0.04
									[0.09, 0.13]	[0.02, 0.05]
										0.05
										[0.04, 0.07]

Note: The numbers in brackets denote the 95% posterior interval.

2.5.5 Predicting CLV at the Category Level

Our proposed model and Bayesian estimation approach allow us to obtain the estimates of customer-specific parameters and thus predict individual customers' purchase patterns associated with each category. This in turn enables us to compute their CLV at the category level and quantify the contribution of individual categories to a firm-level CLV. Prior research has identified various benefits of calculating CLV to develop customer-centric marketing strategies for various purposes such as customer selection, segmentation, and optimal resource allocation (e.g., Reinartz and Kumar 2003, Venkatesan and Kumar 2004). These benefits of the firm-level CLV analysis can be operationalized at the category level by applying our model to category-level transaction data.

Using the estimates of model parameters, we forecast the category-level CLV of the 1,311 customers over three years after the data period.⁷ Figure 2.6 shows the distribution of the metrics across customers for each category. Noteworthy findings are as follows. First, the number of customers who never purchase in a particular product category during the three-year period varies considerably across the categories. For example, the prediction results indicate that 135 (10.3%) out of the 1,311 customers do not purchase creams during the three-year period whereas 460 (35.1%) customers buy no makeups in the time period. Second, with the exception of those customers not purchasing in a product category, the distributions of CLV in the categories of basics, creams, and serums, are right-skewed bell-shaped (with different mean values and standard deviations). In comparison, the distribution of CLV in makeups is relatively flat,

⁷Prior research has employed the revenue stream associated with a customer to compute CLV when information about product costs/margins is not available (e.g., Singh, Borle, and Jain 2009). We follow their approach in this research. We also ignore the time discount factor for cash flows as the time span for our CLV prediction is three years only.

suggesting that the customers are more heterogeneous in their future purchase behavior of the category.⁸ The substantial heterogeneity in customer's expected expenditures across the four product categories suggests there may be value to developing targeted marketing messages.

We then compute the correlations between the category-level CLV across the 1,311 customers to examine how categories are related in terms of their contributions to CLV. Table 2.13 shows the correlation coefficients of the category-level CLV for the six possible pairs of the categories. The results indicate that CLV is significantly correlated between categories except for the pair of basics and makeups.⁹ In particular, we find that the correlation of CLV between creams and serums is much higher than those of other pairs. This implies that customers with a higher CLV in creams (serums) likely have a higher CLV in serums (creams). Thus, in a long run, it may be more effective and beneficial for the firm to induce customers who repeatedly purchase one type of non-basic skin care products (e.g., creams) to buy another type of functional products in a different category (e.g., serums), compared to making cross-selling efforts for other combinations of product categories.

⁸We compute the coefficient of variation, the ratio of the standard deviation to the mean, to compute the dispersion of the distribution of CLV. The coefficient of variation is largest (smallest) for makeups (creams) at 1.08 (0.64).

⁹The t -statistic that tests the significance of the correlation coefficient is given by $\rho \sqrt{\frac{n-2}{1-\rho^2}}$, where ρ is the correlation coefficient and n is the number of observations.

Figure 2.6: Prediction of Three-Year Customer Lifetime Value

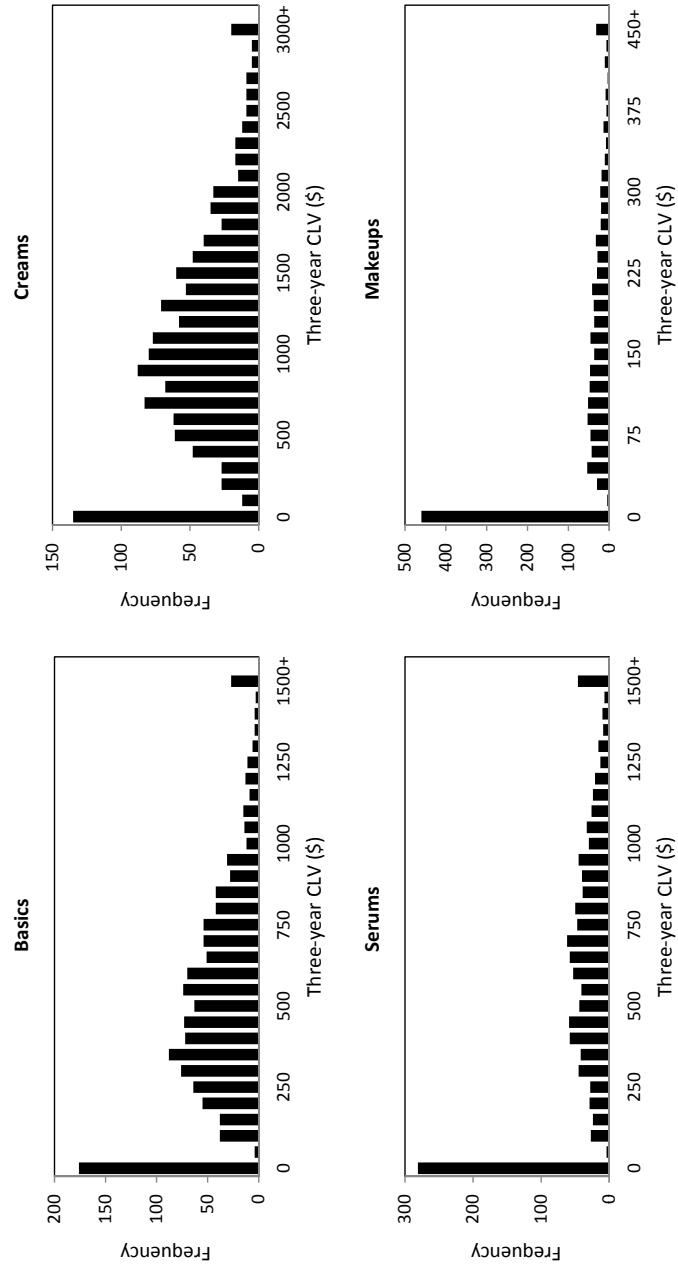


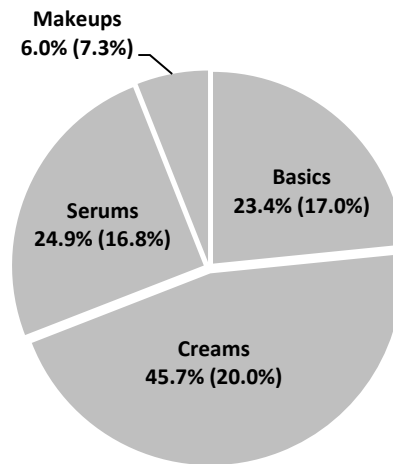
Table 2.13: Correlation of CLV between Categories

	Creams	Serums	Makeups
Basics	0.16**	0.17**	0.04
Creams		0.24**	0.07**
Serums			0.15**

Note: '**' indicate that the correlation coefficient is significant at 0.95 significance level.

A firm-level CLV can be obtained by aggregating the category-level CLV across categories. At the firm level, the three-year CLV of 72 (5.5%) customers are predicted to be zero. The mean CLV per customer is \$2,199 with the standard deviation of \$1,210. To examine the contribution of individual categories to the firm-level CLV, we compute the percentage of the category-level CLV with respect to the firm-level CLV for 1,239 customers whose three-year CLV is non-zero. Figure 2.7 shows the mean value of the percentages averaged across customers for each category and their standard deviation in parentheses. We find that on average, creams account for the largest share of the firm-level CLV with 45.7% and makeups have the least share with 6.0%. We also find a large variation in the percentage contribution of the categories to CLV across customers. In particular, we note that for many customers who have similar CLV at the firm level, their category-level CLV considerably differs from each other. This implies that different customers' lifetime values are driven by different categories, suggesting it may be beneficial to customize marketing offers at the category level as well as at the customer level.

Figure 2.7: Contribution of Individual Categories to CLV



2.5.6 Linking Shopping Basket Choice to CLV

In addition to the category-level analyses, another key benefit of our proposed model is its ability to predict customers' shopping basket choice. The forecast of shopping basket choice can help marketers tailor targeting and promotion campaigns for individual customers by knowing who are more likely to buy a specific combination of categories in the future time period. Furthermore, this allows marketers to relate shopping basket choice to CLV, which could be useful to develop cross-selling strategies.

To provide an illustration, we consider a simple comparison: those customers who purchase a set of categories together at a single transaction vs. those customers who purchase the categories but do so over multiple visits. We compare these two customer groups in terms of their CLV to assess whether purchasing multiple categories together at a single transaction affects CLV. Specifically, we predict the 1,311 customers' purchase patterns across categories over six months after the data period. Then, for each possible set of categories, we

divide customers into two groups: one for customers who have purchased all categories in the set together at least once at the same transaction (Group 1) and one for customers who have purchased each category in the set at least once but never purchased them together at the same transaction (Group 2). We then compute CLV of the customers in each group over the three years after the data period.

Table 2.14 summarizes the prediction results , computed as posterior means across MCMC iterations. For example, the first row of the table indicates that 409 customers purchase basics and creams (and possibly other categories) at least once at the same transaction over the six-month period, and on average their three-year CLV is \$2,284 with the standard deviation of \$1,000. On the other hand, 156 customers purchase both basics and creams in the six-month period but never purchase them together, and their three-year CLV is \$2,262 with the standard deviation of \$946.

Table 2.14: Comparisons of CLV between Customer Groups

Categories	Group 1		Group 2		Difference in CLV
	No. of customers	Three-year CLV	No. of customers	Three-year CLV	
Basics and Creams	409	\$2,284 (\$1,000)	156	\$2,262 (\$946)	\$22
Basics and Serums	355	\$2,312 (\$967)	121	\$2,261 (\$952)	\$51
Basics and Makeups	177	\$2,241 (\$927)	137	\$2,300 (\$849)	\$59
Creams and Serums	417	\$2,427 (\$1,020)	121	\$2,267 (\$903)	\$160*
Creams and Makeups	207	\$2,378 (\$982)	144	\$2,159 (\$905)	\$219**
Serums and Makeups	190	\$2,410 (\$951)	118	\$2,188 (\$896)	\$222**
Basics, Creams, and Serums	217	\$2,335 (\$1,002)	175	\$2,289 (\$965)	\$46
Basics, Creams, and Makeups	97	\$2,338 (\$986)	158	\$2,126 (\$870)	\$212*
Basics, Serums, and Makeups	86	\$2,305 (\$914)	138	\$2,236 (\$864)	\$69
Creams, Serums, and Makeups	112	\$2,433 (\$981)	140	\$2,187 (\$904)	\$246**
Basics, Creams, Serums, and Makeups	53	\$2,475 (\$925)	134	\$2,150 (\$885)	\$325**

Note: (1) The numbers in parentheses denote the standard deviation.

(2) '*' and '**' indicate that the difference in CLV between groups 1 and 2 is significant at 0.90 and 0.95 significance levels, respectively.

From the table, we find that in four out of 15 cases, the three-year average CLV of customers in Group 1 is significantly larger than that of customers in Group 2 at the 95% significance level, and that six out of 15 cases are significant at the 90% significance level. For example, for the case of creams and makeups, customers who purchase both together at least once at the same transaction for the first six months of the prediction period tend to have a higher CLV than those who purchase the two categories separately. The difference in CLV between the two groups is \$219, which is significant at the 95% significance level. For the all cases with significantly different CLV between the two groups, overall we find that the defection rates are higher for customers in Group 1. However, on average they purchase and spend more per transaction than customers in Group 2. When combined, the CLV of customers in Group 1 exceeds that of customers in Group 2, with the effects of larger purchase amounts per transaction outweighing the higher defection rates.

We also find that CLV varies significantly depending on the combination of categories purchased together. For example, customers who purchase basics and creams together at least once at the same transaction have a significantly higher CLV than those who purchase creams and serums together at the same period. To the best of our knowledge, our analysis is the first to relate shopping basket to CLV and provide marketers with guidance in their cross-selling initiatives. The ability of our model to forecast customer purchase patterns across categories enables us to examine the relationships of various other outcomes that may be linked to shopping basket composition.

2.6 Conclusions and Future Research

This research develops an integrated modeling framework for multi-category customer value analysis in a non-contractual setting. We model customers' arrival process to the firm, multi-category purchase incidence and amount decisions, and latent defection in an integrated framework, allowing for shopping dynamics due to the interplay of purchase timing and incidence across categories. To capture the association of shopping basket choice and the arrival process parsimoniously, we propose a novel modeling approach using a latent space of product categories. We also account for the interdependent choices of multiple categories and the correlation of repeatedly observed outcomes both within and across categories.

Applying our proposed model to category-level customer transaction data from a leading beauty care company in Korea, we find that our model offers excellent fit and performance in predicting customer purchase patterns across categories. This enables us to measure CLV at the category level and assess the contribution of each category to a firm-level CLV. Our results also reveal that customers who purchase multiple categories together at the same transaction tend to have a higher CLV compared to those purchasing the same categories separately over multiple visits. Given this finding, our research provides a framework within which the impact of promotions designed to enhance cross-selling can be evaluated in terms of their impact on future expenditures.

There are a number of limitations that should be acknowledged and perhaps addressed in future research. First, given the purpose of this research and the limitation of the data, we have not considered the impact of marketing mix

variables on customer shopping behavior. Incorporating such variables into the various modeling components is straightforward, and the addition of an optimization layer to the model would allow us to examine how marketing actions by a firm influence purchase patterns across categories. The resulting inferences could be useful to effectively manage customer relationship and driving CLV. While we do not have access to such data in our empirical application, our framework provides a general platform for the inclusion of covariates. For example, one could easily add more covariates into the timing model of customer arrivals in equation (2.2), the multi-category choice model in equations (2.3) and (2.4), the amount model in equations (2.7) and (2.8). We also note that our timing model can be also modified to bring in the methods proposed by Gupta (1991), who demonstrated a sophisticated way of bringing time-varying covariates into a multi-event timing model.

Second, this research has focused on studying customer purchase patterns in the offline channel. Given the growing popularity of e-commerce, an increasing number of customers make transactions with a firm through both its online and offline channels (e.g., Ansari, Mela, and Neslin 2008). Along this line, restricting the analysis to a single channel can leave out important information regarding customer retention and attrition. Another area for future research is therefore to incorporate the role of different channels and their interactive effect into our proposed model for CLV analysis in multi-category and multi-channel contexts. We hope this study generates further interest and accelerate the progress in this important area.

Appendix: Derivation of Equation (2.5)

Let us denote the probability of the “no-purchase” case derived using equation (2.3) as

$$\Pr(C_{ij1}^0, C_{ij2}^0, \dots, C_{ijK}^0) = \Pr(C_{ij1} = 0, C_{ij2} = 0, \dots, C_{ijK} = 0). \quad (A1)$$

Then, from the theorem by Besag (1974), we have

$$\begin{aligned} \frac{\Pr(C_{ij1}, C_{ij2}, \dots, C_{ijK})}{\Pr(C_{ij1}^0, C_{ij2}^0, \dots, C_{ijK}^0)} &= \frac{\Pr(C_{ij1}|C_{ij2}, C_{ij3}, \dots, C_{ijK})}{\Pr(C_{ij1}^0|C_{ij2}, C_{ij3}, \dots, C_{ijK})} \cdot \frac{\Pr(C_{ij2}|C_{ij1}^0, C_{ij3}, \dots, C_{ijK})}{\Pr(C_{ij2}^0|C_{ij1}^0, C_{ij3}, \dots, C_{ijK})} \\ &\quad \cdots \frac{\Pr(C_{ijK}|C_{ij1}^0, C_{ij2}^0, \dots, C_{ij,K-1}^0)}{\Pr(C_{ijK}^0|C_{ij1}^0, C_{ij2}^0, \dots, C_{ij,K-1}^0)}. \end{aligned} \quad (A2)$$

By substituting equation (2.3) into equation (A2), we have

$$\begin{aligned} \Pr(C_{ij1}, C_{ij2}, \dots, C_{ijK}) &= \Pr(C_{ij1}^0, C_{ij2}^0, \dots, C_{ijK}^0) \cdot \{\exp(\pi_{ij1} + \Theta_1 \mathbf{C}_{ij} + \Psi_1 \mathbf{C}_{i,j-1})\}^{C_{ij1}} \\ &\quad \cdot \{\exp(\pi_{ij2} + \Theta_2 \mathbf{C}_{ij}^{(023\dots K)} + \Psi_2 \mathbf{C}_{i,j-1})\}^{C_{ij2}} \\ &\quad \cdots \{\exp(\pi_{ijK} + \Theta_K \mathbf{C}_{ij}^{(000\dots K)} + \Psi_K \mathbf{C}_{i,j-1})\}^{C_{ijK}}, \end{aligned} \quad (A3)$$

where Θ_k and Ψ_k are the k th row vector of matrices Θ and Ψ , respectively, and

$$\mathbf{C}_{ij}^{(023\dots K)} = (0, C_{ij2}, C_{ij3}, \dots, C_{ijK}) \text{ and } \mathbf{C}_{ij}^{(000\dots K)} = (0, 0, 0, \dots, C_{ijK}).$$

Equation (A3) can be written as:

$$\begin{aligned} \Pr(C_{ij1}, C_{ij2}, \dots, C_{ijK}) &= \Pr(C_{ij1}^0, C_{ij2}^0, \dots, C_{ijK}^0) \cdot \exp\{C_{ij1}(\pi_{ij1} + \Theta_1 \mathbf{C}_{ij} + \Psi_1 \mathbf{C}_{i,j-1})\} \\ &\quad \cdot \exp\{C_{ij2}(\pi_{ij2} + \Theta_2 \mathbf{C}_{ij}^{(023\dots K)} + \Psi_2 \mathbf{C}_{i,j-1})\} \\ &\quad \cdots \exp\{C_{ijK}(\pi_{ijK} + \Theta_K \mathbf{C}_{ij}^{(000\dots K)} + \Psi_K \mathbf{C}_{i,j-1})\}. \end{aligned} \quad (A4)$$

By arranging exponential terms in equation (A4), we have

$$\begin{aligned} \Pr(C_{ij1}, C_{ij2}, \dots, C_{ijK}) &= \Pr(C_{ij1}^0, C_{ij2}^0, \dots, C_{ijK}^0) \\ &\quad \cdot \exp(\pi_{ij}' \mathbf{C}_{ij} + \frac{1}{2} \mathbf{C}_{ij}' \Theta \mathbf{C}_{ij} + \mathbf{C}_{i,j-1}' \Psi \mathbf{C}_{i,j-1}). \end{aligned} \quad (A5)$$

Because the sum of the joint probability in equation (A5) across all possible combinations of categories should equal 1 and our data do not include the “no-purchase” case, the joint distribution of $\mathbf{C}_{ij} = \{C_{ij1}, C_{ij2}, \dots, C_{ijK}\}'$ is given by:

$$P\left(\mathbf{C}_{ij} = \mathbf{C}_{ij}^*, \sum_{k=1}^K C_{ijk}^* \geq 1\right) = \frac{\exp(\boldsymbol{\pi}_{ij}' \mathbf{C}_{ij}^* + \frac{1}{2} \mathbf{C}_{ij}^{*'} \boldsymbol{\Theta} \mathbf{C}_{ij}^* + \mathbf{C}_{i,j-1}' \boldsymbol{\Psi} \mathbf{C}_{i,j-1})}{\sum_{\mathbf{C}_{ij} \neq \{0,0,\dots,0\}} \exp(\boldsymbol{\pi}_{ij}' \mathbf{C}_{ij} + \frac{1}{2} \mathbf{C}_{ij}' \boldsymbol{\Theta} \mathbf{C}_{ij} + \mathbf{C}_{i,j-1}' \boldsymbol{\Psi} \mathbf{C}_{i,j-1})}. \quad (\text{A6})$$

References

- Abe, Makoto (2009), "Counting Your Customers One by One: A Hierarchical Bayes Extension to the Pareto/NBD Model," *Marketing Science*, 28 (3), 541-553.
- Ansari, Asim, Carl F. Mela, and Scott A. Neslin (2008) "Customer Channel Migration," *Journal of Marketing Research*, 45 (1), 60-76.
- Bell, David, John Deighton, Werner J. Reinartz, Roland T. Rust, and Gordon Swartz (2002) "Seven Barriers to Customer Equity Management," *Journal of Service Research*, 5 (1), 77-86.
- Besag, Julian (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society: Series B*, 36 (2), 192-236.
- Bradlow, Eric, Bart Bronnenberg, Gary Russell, Neeraj Arora, David Bell, Sri Duvvuri, Frankel Hofstede, Catarina Sismeiro, Raphael Thomadsen, and Sha Yang (2005), "Spatial Models in Marketing," *Marketing Letters*, 16 (3), 267-278.
- Bradlow, Eric and David C. Schmittlein (2000), "The Little Engines That Could: Modeling the Performance of World Wide Web Search Engines," *Marketing Science*, 19 (1), 43-62.
- Chib, Siddhartha, P. B. Seetharaman, and Andrei Strijnev (2002), "Analysis of Multi-Category Purchase Incidence Decisions Using IRI Market Basket Data," *Advances in Econometrics*, 16, 57-92.
- Chintagunta, Pradeep K. (1999), "Variety Seeking, Purchase Timing, and the 'Lightning Bolt' Brand Choice Model," *Management Science*, 45 (4), 468-498.
- Chintagunta, Pradeep K. and Sudeep Haldar (1998), "Investigating Purchase Timing Behavior in Two Related Product Categories," *Journal of Marketing Research*, 35 (1), 43-53.
- DeSarbo, Wayne S. and Jianan Wu (2001), "The Joint Spatial Representation of Multiple Variable Batteries Collected in Marketing Research," *Journal of Marketing Research*, 38 (2), 244-253.
- Ehrenberg, A. S. C. (1988), *Repeat-Buying: Facts, Theory and Applications*, 2nd edn., London: Griffin; New York: Oxford University Press.
- Fader, Peter S. and Bruce G.S. Hardie (2009), "Probability Models for Customer-Base Analysis," *Journal of Interactive Marketing*, 23 (1), 61-69.

- Fader, Peter S., Bruce G. S. Hardie, and Chun-Yao Huang (2004), "A Dynamic Change-point Model for New Product Sales Forecasting," *Marketing Science*, 23 (3), 280-303.
- Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005a), "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis," *Journal of Marketing Research*, 42 (4), 415-430.
- Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005b), "Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model," *Marketing Science*, 24 (2), 275-284.
- Fader, Peter S. and Leonard M. Lodish (1990), "A Cross-Category Analysis of Category Structure and Promotional Activity for Grocery Products," *Journal of Marketing*, 54 (4), 52-65.
- Gelfand, Alan E. and Adrian F. M. Smith (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85 (410), 398-409.
- Gupta, Sunil (1988), "Impact of Sales Promotions on When, What, and How Much to Buy," *Journal of Marketing Research*, 25 (4), 342-355.
- Gupta, Sunil (1991), "Stochastic Models of Interpurchase Time with Time-Dependent Covariates," *Journal of Marketing Research*, 28 (1), 1-15.
- Gupta, Sunil, Dominique Hanssens, Bruce Hardie, William Kahn, V. Kumar, Nathaniel Lin, Nalini Ravishanker, and S. Sriram (2006), "Modeling Customer Lifetime Value," *Journal of Service Research*, 9 (2), 139-155.
- Gupta, Sunil and Donald R. Lehmann, and Jennifer Ames Stuart (2004), "Valuing Customers," *Journal of Marketing Research*, 41 (1), 7-18.
- Hastings, W.K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57 (1), 97-109.
- Jen, Lichung, Chien-Heng Chou, and Greg M. Allenby (2009), "The Importance of Modeling Temporal Dependence of Timing and Quantity in Direct Marketing," *Journal of Marketing Research*, 46 (4), 482-493.
- Kamakura, Wagner A., Sridhar Ramaswami, and Rajenda K. Srivastava (1991), "Applying Latent Trait Analysis in the Evaluation of Prospects for Cross-Selling of Financial Services," *International Journal of Research in Marketing*, 8 (4), 329-349.
- Kumar, V. (2008), *Managing Customers for Profit: Strategies to Increase Profits and Build Loyalty*, 1st edn., Pearson Prentice Hall.

- Kumar, V., Rajkumar Venkatesan, and Werner Reinartz (2008), "Performance Implications of Adopting a Customer-Focused Sales Campaign," *Journal of Marketing*, 72 (5), 50-68.
- Li, Shibo, Baohong Sun, and Ronald T. Wilcox (2005), "Cross-Selling Sequentially Ordered Products: An Application to Consumer Banking Services," *Journal of Marketing Research*, 42 (2), 233-239.
- Manchanda, Puneet, Asim Ansari, and Sunil Gupta (1999), "The 'Shopping Basket': A Model for Multicategory Purchase Incidence Decisions," *Marketing Science*, 18 (2), 95-114.
- Mehta, Nitin (2007), "Investigating Consumers' Purchase Incidence and Brand Choice Decisions Across Multiple Product Categories: A Theoretical and Empirical Analysis," *Marketing Science*, 26 (2), 196-217.
- Moe, Wendy W. and Peter S. Fader (2004), "Capturing Evolving Visit Behavior in Clickstream Data," *Journal of Interactive Marketing*, 18 (1), 5-19.
- Moon, Sangkil and Gary J. Russell (2008), "Predicting Product Purchase from Inferred Customer Similarity: An Autologistic Model Approach," *Marketing Science*, 54 (1), 71-82.
- Morrison, Donald G. and David C. Schmittlein (1988), "Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is It Worth the Effort?," *Journal of Business and Economic Statistics*, 6 (2), 145-159.
- Narasimhan, Chakravarthi, Scott A. Neslin, and Subrata K. Sen (1996), "Promotional Elasticities and Category Characteristics," *Journal of Marketing*, 60 (2), 17-30.
- Niraj, Rakesh, V. Padmanabhan, and P. B. Seetharaman (2008), "A Cross-Category Model of Households' Incidence and Quantity Decisions," *Marketing Science*, 27 (2), 225-235.
- Park, Young-Hoon and Peter S. Fader (2004), "Modeling Browsing Behavior at Multiple Websites," *Marketing Science*, 23 (3), 280-303.
- Reinartz, Werner J. and V. Kumar (2000), "On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing," *Journal of Marketing*, 64 (4), 17-35.
- Reinartz, Werner J. and V. Kumar (2003), "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of Marketing*, 67 (1), 77-99.

- Russell, Gary J. and Ann Petersen (2000), "Analysis of Cross Category Dependence in Market Basket Selection," *Journal of Retailing*, 76 (3), 367-392.
- Schmittlein, David C., Donald G. Morrison, and Richard Colombo (1987), "Counting Your Customers: Who Are They and What Will They Do Next?," *Management Science*, 33 (1), 1-24.
- Schmittlein, David C. and Robert A. Peterson (1994), "Customer Base Analysis: An Industrial Purchase Process Application," *Marketing Science*, 13 (1), 41-67.
- Schweidel, David A., Peter S. Fader, and Eric T. Bradlow (2008), "A Bivariate Timing Model of Customer Acquisition and Retention," *Marketing Science*, 27 (5), 829-843.
- Singh, Siddharth S., Sharad Borle, and Dipak C. Jain (2009), "A Generalized Framework for Estimating Customer Lifetime Value When Customer Lifetimes are Not Observed," *Quantitative Marketing and Economics*, 7 (2), 181-205.
- Song, Inseong and Pradeep K. Chintagunta (2006), "Measuring Cross-Category Price Effects with Aggregate Store Data," *Management Science*, 52 (10), 1594-1609.
- Venkatesan, Rajkumar and V. Kumar (2004), "A Customer Lifetime Value Framework for Customer Selection and Optimal Resource Allocation Strategy," *Journal of Marketing*, 68 (4), 106-125.
- Wagner, Udo and Alfred Taudes (1986), "A Multivariate Polya Model of Brand Choice and Purchase Incidence," *Marketing Science*, 5 (3), 219-244.

CHAPTER 3

MODELING ONLINE VISITATION AND CONVERSION DYNAMICS

3.1 Introduction

Ever since the Internet emerged as a medium for commercial purpose, online shopping has seen phenomenal growth. In the United States, with a 10% compound annual growth rate, online retail sales is expected to reach 250 billion dollars by 2014 from 155 billion dollars in 2009 (*Forrester Research* 2010). Despite the success of e-commerce channel, however, most online retailers have been able to convert less than 5% of customer visits into purchases. Naturally, understanding conversion behavior has become of great importance to online retailers because small changes in conversion rates can lead to considerable increases in revenues.

The low rate of visit-to-purchase conversion has also motivated academic research on online purchase behavior (e.g., Hauser et al. 2009, Montgomery et al. 2004, Sismeiro and Bucklin 2004). Many studies in this stream of research have examined conversion behavior by taking into account the activities taken or the pages viewed by a customer *within* her visit to an online retailer. Distinctly from others, Moe and Fader (2004) model conversion behavior *across* store visits with an assumption that a customer's purchase likelihood is reset to a baseline value at every purchase occasion and increases as she makes more visits. Yet little research considers online shopping patterns from a timing perspective to improve our understanding of how a series of store visits lead to purchase conversions. In this research, we develop a dynamic model to capture the store visit patterns of online customers and link them with conversion behavior.

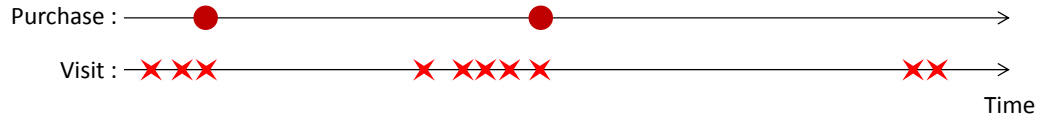
Behavioral research on customer buying process shows that a customer's purchase decision is preceded by the stages of information search and evaluation (e.g., Moorthy, Ratchford, and Talukdar 1997, Urbany, Dickson, and Wilkie 1989). The intensity of such preparatory tasks varies across individuals depending on the cost of conducting the activities and the potential benefits/regrets associated with the purchase decision. For example, the purchase of a high involvement product often involves a long period of information search. Customers with lower search costs collect and evaluate more information prior to a purchase.

In online shopping, the low transportation (browsing) costs required to visit an e-commerce site encourage customers to make several shopping trips to the online store for the prepurchase activities before making a buying decision (e.g., Johnson et al. 2004, Moe and Fader 2004). This implies that longitudinal customer shopping behavior at an e-commerce site can be characterized by occasional shopping goals (i.e., purchases of items in needs) and a series of frequent visits to attain each shopping goal. Therefore, a customer's overall visit process to the online store tends to consist of multiple visit clusters.

Consider an example of customer shopping patterns chosen from the database we use for this research. Figure 3.1 illustrates a sequence of 10 visits and two purchase conversions by a customer of an online store over a period of 60 days. The customer's arrival patterns can be described as a point event process on a time dimension, which is broadly classified as random, uniform or clustered (Dacey and Tung 1962). At first glance, we note that the customer's visits to the website tend to occur in a clustered manner rather than randomly: The customer made three successive visits at the beginning of the time period,

and became temporarily inactive for a few weeks. She then returned to the retailer and made several visits thereafter before another long hiatus. Toward the end of the 60-day period, she made a couple of visits to the website, again clustered together. The 10 visits by this customer therefore can be partitioned into multiple visit clusters, based on the relative temporal proximity between the events. We also find various patterns of clustered visits by other customers in our database.

Figure 3.1: Lumpy Shopping Patterns of an Online Customer



Using the existing multi-event timing models in the marketing literature (e.g., Fader, Hardie, and Lee 2005, Gupta 1991, Schmittlein, Morrison, and Colombo 1987), marketers can forecast future shopping patterns by the customer illustrated in Figure 3.1. A key question arises here as to whether these predictions could be improved by taking into account the lumpiness embedded in the customer's visit process. The preceding example suggests the possibility of improvements given the existence of the customer's recurring visit clusters. For example, after the last visit in the data period, the future intervisit time is likely to be smaller if her next visit belongs to the last visit cluster, compared to the case in which the next visit initiates a new visit cluster.

Assuming that the online customer's visits can be divided into multiple visit clusters, another important question for marketers of the online store would be how the inferred formation of visit clusters can be leveraged to predict the likelihood of purchase conversion upon a customer visit. When a customer makes

multiple frequent visits to an online store for a purchase decision, the shopping occasion could be better represented by the corresponding visit cluster rather than individual store visits, and as a result the conversion probability may change systematically within the visit cluster. Therefore, the inferences based on visit clusters could be also useful in understanding online purchase behavior.

We develop a model to examine the clustered visit patterns of online customers and their purchase behavior across store visits. Our model is based on the notion that the arrival process of customer visits tends to consist of multiple visit clusters with a relatively high visit rate within a cluster and a smaller visit rate between clusters. To capture this nonstationarity in the arrival process, we assume that a customer's visit rate varies over time depending on the formation of visit clusters, while allowing the probability of making clustered visits to change dynamically as she progresses through multiple visits within a cluster. Given the latency of visit clusters, we take a changepoint modeling approach and statistically infer the cluster formation on the basis of the customer's visit patterns through data augmentation in Bayesian approach. The model also accounts for customer defections and various sources of customer heterogeneity in a flexible manner.

Our proposed model thereby offers not only better predictions of customer visit patterns and purchase conversions but also more in-depth understanding of online shopping behavior, beyond just reporting overall counts of visits by individual customers. As a key benefit, our model provides a set of novel inferences about the patterns that underlie the shopping behavior of online customers. Specifically, it allows us to infer (1) the intervisit time within a cluster,

(2) the number of visits per cluster, (3) the time length of a cluster, (4) the inter-visit time between clusters, and (5) the number of visit clusters in a given time period, at the individual customer level.

We apply our model to the database of customer visits and purchases at a major online retailer in the United States. We find strong empirical evidence of lumpy visit patterns by online customers with significant heterogeneity in the extent of the lumpiness. As a result, our model offers excellent fit and predictive performance in a comparison with extant multi-event timing models. As part of our substantive contribution, we highlight how the inferred formation of visit clusters based on store visit patterns can be leveraged to examine conversion behavior of online customers. We find that a customer's purchase likelihood changes across visits depending on how many visits the customer has made previously within a visit cluster. By combining the cluster-based inferences with the data of the reference sites to the online store, we also find that the conversion rate varies considerably depending on the sequence of reference sites chosen by customers.

From the methodological perspective, this research is in line with a research stream on understanding and predicting customers' visit or purchase timing behavior, a key objective of marketing research for years. One of the standard modeling frameworks is the negative binomial distribution (NBD) model which characterizes customers' interarrival times by an exponential distribution and heterogeneity in their time-invariant mean arrival rates by a gamma distribution (e.g., Morrison and Schmittlein 1981, 1988). Several researchers have extended the NBD model to incorporate unobserved customer defections (e.g., Fader, Hardie, and Lee 2005, Schmittlein, Morrison, and Colombo 1987, Singh,

Borle, and Jain 2009), nonstationarity in the customer arrival process (e.g., Fader, Hardie, and Huang 2004), interdependence between two correlated processes (e.g., Park and Fader 2004, Schweidel, Fader, and Bradlow 2008), and the effect of time-varying explanatory variables (e.g., Fader, Hardie, and Huang 2004, Gupta 1991). The NBD-based models are not the only way to capture customer shopping patterns and another popular modeling approach is the proportional hazard model (PHM), where the hazard rate is given by the product of the baseline hazard function and the covariate function. Several researchers employ or extend the PHM to study stochastic interarrival or interpurchase times (e.g., Chintagunta and Haldar 1998, Jain and Vilcassim 1991, Seetharaman and Chintagunta 2003, Telang, Boatwright, and Mukhopadhyay 2004, Wedel et al. 1995).

The contribution of our research is twofold. First, we develop a novel multi-event timing model of customer visits that explicitly considers nonstationarity due to recurring visit clusters in the arrival process. Our model is flexible and general in that it nests several established timing models. Given the unique aspect of online shopping behavior, we show that our model outperforms the existing timing models. Because our model features a general structure, it can be also applied to various other shopping contexts where customer visits (or transactions) occur in a clustered manner. Second, our proposed model enables us to infer the formation of latent visit clusters and obtain useful inferences about the online shopping patterns of individual customers. In particular, given that visit clusters could serve as a reasonable proxy for sales opportunities the online retailer had, the inferred cluster formation allows marketers to tract which customer visits have more influences on sales by relating a purchase conversion with store visits occurred in the same cluster. Hence, marketers could use the results to evaluate the effectiveness of their online campaigns, and allocate more

resources to successful marketing actions associated with visits which lead to purchase conversions.

The remainder of this chapter is organized as follows. In section 3.2, we provide a detailed specification of our model. Section 3.3 describes the data used in our empirical analysis. In section 3.4, we discuss the model results and illustrate insights afforded by analyzing online customers' lumpy visit patterns. Finally, section 3.5 concludes and suggests future research directions.

3.2 Model

Our modeling objective is to capture the lumpy visit patterns of online customers and their purchase behavior across store visits. We first build our model and discuss how to formulate the likelihood function. We then compare and contrast our proposed model with several benchmark models to highlight its key properties. This is followed by the discussion on our computational approach to estimating the model and a simulation study to demonstrate the efficacy of the model and estimation method.

3.2.1 Model Structure

During the period $[0, T]$, where 0 corresponds to the beginning of the data period and T is the censoring point that corresponds to the end of the model calibration period, we observe customer i making J_i visits to the online store at times $t_{i1}, t_{i2}, \dots, t_{iJ_i}$. We also observe whether the customer makes a purchase or not at each visit, and denote the purchase incidence decision of customer i at

her j th visit as Y_{ij} .

We model the two sets of sequential behavioral outcomes with respect to the visit process for the online customer and her purchase decision at each visit. Given that purchase events are always conditional on store visits, the overall modeling framework (for customer i) can be represented as

$$\Pr(\mathbf{T}_i, \mathbf{Y}_i) = \Pr(\mathbf{Y}_i | \mathbf{T}_i) \Pr(\mathbf{T}_i), \quad (3.1)$$

where $\mathbf{T}_i = (t_{i1}, t_{i2}, \dots, t_{iJ_i})$ and $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iJ_i})$. We thereby provide a formal layout of the model to capture the customer's purchase behavior conditional on her visit (i.e., $\mathbf{Y}_i | \mathbf{T}_i$) in section 3.2.2 and the underlying store visit process (i.e., \mathbf{T}_i) in section 3.2.3.

3.2.2 Purchase Model

We model customer i 's purchase incidence decision at her j th visit to the online store. The customer's purchase behavior at a given visit may be predictable based on various factors including her within-visit browsing activities (e.g., the number of webpages viewed by the customer during the visit, the time duration of the visit), the timing of the visit (e.g., weekend, holiday shopping seasons), and her purchase history (e.g., the purchase decisions at prior visits). In addition, given the online customer's tendency of making multiple visits for a single purchase decision, we expect that the likelihood of making a purchase may change as the customer progresses through multiple shopping trips to the store. We also expect that when multiple visits are made for the same shopping goal, the visits would be temporally close to each other, thus forming a cluster of visits. Taken together, this implies that the purchase probability may vary

across visits within a visit cluster.¹

To capture this process, we specify customer i 's purchase probability at her j th visit, using the following logit model:

$$P(Y_{ij} = 1) = \frac{\exp\{\alpha_{0i} + \alpha_1 \mathbf{X}_{ij} + f(K_{ij})\}}{1 + \exp\{\alpha_{0i} + \alpha_1 \mathbf{X}_{ij} + f(K_{ij})\}}, \quad (3.2)$$

where α_{0i} captures individual-specific characteristics affecting the purchase likelihood and follows $N(\mu_\alpha, \sigma_\alpha^2)$; \mathbf{X}_{ij} is a vector of individual-specific time-varying covariates for customer i at the j th visit; K_{ij} is the cumulative number of visits made by customer i in the current visit cluster up to her j th visit; $f(\cdot)$ is a function of K_{ij} to be specified in section 3.4.1.

At the heart of our purchase model is the variable K_{ij} employed to take into account the multi-visit tendency for a single purchase decision online. By definition, K_{ij} is set to 1 at the beginning of each cluster and increases as customer i makes additional visits within a cluster. The values of K_{ij} are thus determined through the formation of visit clusters, or reversely the cluster formation for customer i can be represented using $\mathbf{K}_i = (K_{i1}, K_{i2}, \dots, K_{iJ_i})$. It is however important to note that what we observe in the data is customer visits not visit clusters, and oftentimes the formation of visit clusters is not apparent. Hence, without the information on the cluster formation on hand, the values of K_{ij} are not readily available. We therefore infer the formation of visit clusters (and thus the values of K_{ij}), using the timing model of customer visits presented in the next section.

Finally, to obtain the likelihood function of the purchase model (conditional

¹Given that Internet clickstream data can capture detailed information about browsing behavior of individuals, an alternative way of counting store visits made for a purchase decision would be to investigate the contents of webpages or URLs viewed by a customer at her each visit to the online store. However, as Moe and Fader (2004) point out, in practice, it is very costly and difficult to maintain and manipulate such sizable databases. Moreover, this approach could be still misleading, because online customers often view many webpages or products including ones irrelevant to their shopping purpose.

on \mathbf{K}_i), we multiply the purchase probability in equation (3.2) across customers while integrating over the individual-specific intercept α_{0i} :

$$L_{\text{Purchase}} | \mathbf{K}_i = \prod_{i=1}^N \prod_{j=1}^{J_i} \left[\int P(Y_{ij} = 1)^{\mathbb{I}[Y_{ij}=1]} \{1 - P(Y_{ij} = 1)\}^{\mathbb{I}[Y_{ij}=0]} dF(\alpha_{0i}) \right], \quad (3.3)$$

where N is the number of customers in the calibration data and the super-scripted indicator function $\mathbb{I}[\cdot]$ equals 1 if the expression is true and 0 otherwise.

3.2.3 Visit Model

To model customer i 's store visit process, we specify the timing model for the set of $J_i - 1$ intervisit times, $t_{i2} - t_{i1}, t_{i3} - t_{i2}, \dots, t_{iJ_i} - t_{i,J_i-1}$, and the right-censored observation $T - t_{iJ_i}$. We assume that customer i 's j th intervisit time follows an exponential distribution with visit rate λ_{ij} . Then, the density function for the intervisit times and the survival function for the right-censored observation are given by:

$$f(t_{i,j+1} | t_{ij}; \lambda_{ij}) = \lambda_{ij} \exp\{-\lambda_{ij}(t_{i,j+1} - t_{ij})\} \text{ and } S(T | t_{iJ_i}; \lambda_{iJ_i}) = \exp\{-\lambda_{iJ_i}(T - t_{iJ_i})\}. \quad (3.4)$$

The exponential assumption has been widely adopted in the marketing literature because of its parsimony and performance (e.g., Fader, Hardie, and Lee 2005, Park and Fader 2004, Schmittlein, Morrison, and Colombo 1987).

As discussed, the lumpy visit patterns online imply that the visit process of an online shopper can be characterized by multiple clusters of visits with relatively short intervisit times within each cluster and longer intervisit times between clusters (i.e., between the last visit in one cluster and the first visit in the subsequent cluster). Formally, this suggests that the visit frequency within a visit cluster tends to be greater than that between clusters. To take into account

this nonstationarity in the visit process, we assume the visit rate for customer i 's j th intervisit time is given by:

$$\lambda_{ij} = \begin{cases} \lambda_i^w & \text{if customer } i \text{ makes a visit within the current cluster,} \\ \lambda_i^b (< \lambda_i^w) & \text{otherwise.} \end{cases} \quad (3.5)$$

Our next step is to model the formation of latent visit clusters in the visit process. We assume that after the j th visit to the online store, customer i stays within the current cluster with probability p_{ij} and leaves that cluster with probability $(1 - p_{ij})$. Therefore, combined with equation (3.5), this implies that the intervisit time between the j th and $(j + 1)$ th visits is driven by the visit rate of λ_i^w with probability p_{ij} and by the rate of λ_i^b with probability $(1 - p_{ij})$.

Similarly with the purchase probability, the customer's likelihood of making clustered visits may vary based on various factors we can observe. In addition, if clustered visits occur because of the multi-visit tendency for a single purchase decision online, the probability of making another visit in the current cluster would also depend on how many store visits she previously made for the purchase decision. Accordingly, we model p_{ij} as:

$$p_{ij} = \frac{\exp\{\beta_{0i} + \beta_1 \mathbf{Z}_{ij} + g(K_{ij})\}}{1 + \exp\{\beta_{0i} + \beta_1 \mathbf{Z}_{ij} + g(K_{ij})\}}, \quad (3.6)$$

where β_{0i} represents individual-specific characteristics affecting the formation of latent visit clusters; \mathbf{Z}_{ij} is a vector of individual-specific time-varying covariates for customer i 's j th visit; as defined in equation (3.2), the latent variable K_{ij} is the number of visits made by customer i in the current visit cluster; $g(\cdot)$ is a function of K_{ij} to be specified in section 3.4.1.

In our noncontractual context, customers do not notify the online store if they stop shopping at the site; instead, they just silently defect. Accordingly,

the failure to consider the unobserved customer dropouts can result in biased estimates of the visit rates. To account for customer defections, we assume that after any visit, customer i becomes permanently inactive with probability q_i , following the beta-geometric/NBD (BG/NBD) model by Fader, Hardie, and Lee (2005). Therefore, the point at which the customer drops out is distributed across visits, according to a (shifted) geometric distribution with probability density:

$$P(\text{customer } i \text{ drops out after her } j\text{th visit}) = q_i(1 - q_i)^{j-1}. \quad (3.7)$$

In modeling a sequence of customer visits, we expect that customers are heterogeneous in not only their visit frequency and defections but also their tendency for the formation of latent visit clusters. To incorporate customer heterogeneity into our model, we specify the model parameters as follows. First, for the state-specific visit rates, we assume that $\lambda_i^w = \lambda_i^b + \delta_i$, where λ_i^b and δ_i follow a lognormal distribution. This ensures that λ_i^w is greater than λ_i^b . Second, we assume the parameter β_{0i} in equation (3.6) follows a normal distribution. Third, we reparameterize the dropout probability q_i as $q_i = \frac{\exp(\omega_i)}{1+\exp(\omega_i)}$ to ensure $q_i \in [0, 1]$, and assume that ω_i follows a normal distribution. Taken together, we assume

$$\begin{bmatrix} \log \lambda_i^b \\ \log \delta_i \\ \beta_{0i} \\ \omega_i \end{bmatrix} \sim \text{MVN} \left(\begin{bmatrix} \mu_\lambda \\ \mu_\delta \\ \mu_\beta \\ \mu_\omega \end{bmatrix}, \Sigma \right), \quad (3.8)$$

to allow interdependence among the model parameters.

Likelihood Function

The proposed timing model allows us to describe the lumpy visit patterns of online customers while capturing ongoing dynamics within a visit cluster. How-

ever, the benefit comes at a cost. Because we do not observe at which visit each cluster begins and ends, the values of K_{ij} are not readily available. Without the information of visit clusters, there are 2^{J_i-1} possible ways of clustering customer i 's J_i visits. Moreover, there are three possible scenarios for the right-censored time period, $T - t_{iJ_i}$, according to whether or not the customer has dropped out and which visit rate drives the censored observation (if the customer is alive). With this complexity embedded in our model, we employ a changepoint modeling framework to formulate the likelihood function of the model.²

Suppose there are $m_i (\leq J_i)$ visit clusters for customer i 's J_i visits during the data period. We denote the latent set of visit clusters as:

$$\mathbf{\Gamma}_i = \{ \underbrace{(u_{i1}, v_{i1})}_{\text{1st cluster}}, \underbrace{(u_{i2}, v_{i2})}_{\text{2nd cluster}}, \dots, \underbrace{(u_{im_i}, v_{im_i})}_{m_i\text{th cluster}} \}, \quad (3.9)$$

where u_{ic} and v_{ic} indicate the first and last visits in cluster c , respectively. Note that the last cluster is only partially observed, because the data on customer visits are right-censored. Thus, we categorize the last observed visit in the last cluster, denoted as v_{im_i} , into the following three cases: (1) E1 if customer i permanently drops out after the J_i th visit, (2) E2 if the customer is alive and the last cluster ends with the J_i th visit, and (3) E3 if the customer is alive and the last cluster does not end after the J_i th visit. Because the visit rate within a cluster (λ_i^w) is different from the visit rate between clusters (λ_i^b), there are $(2m_i - 2)$ changepoints (at $v_{i1}, u_{i2}, v_{i2}, u_{i3}, \dots, v_{i,m_i-1}, u_{im_i}$) if $v_{im_i} = \text{E3}$, and otherwise, there are $(2m_i - 1)$ changepoints (at $v_{i1}, u_{i2}, v_{i2}, u_{i3}, \dots, u_{im_i}, v_{im_i}$).

In the example illustrated in Figure 3.1, let us assume that the 10 visits by the customer are partitioned into three clusters, one cluster with the first three

²We refer readers to Barry and Hartigan (1993) and Pievatolo and Rotondi (2000) for detailed treatments of the general changepoint model.

visits, the second cluster with the next five visits, and the third cluster with the last two visits. If we further assume that the last visit cluster ends with the 10th visit and the customer is alive after the last visit, the corresponding cluster formation can be represented as $\Gamma_i = \{(1, 3), (4, 8), (9, E2)\}$ with five changepoints. Then, using p_{ij} and q_i , the probability of the cluster formation for the customer is given by:

$$\begin{aligned}
P(\Gamma_i) = & \underbrace{(1 - q_i)p_{i1} \cdot (1 - q_i)p_{i2} \cdot (1 - q_i)(1 - p_{i3})}_{\text{1st cluster}} \\
& \cdot \underbrace{(1 - q_i)p_{i4} \cdot (1 - q_i)p_{i5} \cdot (1 - q_i)p_{i6} \cdot (1 - q_i)p_{i7} \cdot (1 - q_i)(1 - p_{i8})}_{\text{2nd cluster}} \\
& \cdot \underbrace{(1 - q_i)p_{i9} \cdot (1 - q_i)(1 - p_{i10})}_{\text{3rd cluster}}, \tag{3.10}
\end{aligned}$$

where, for example, the likelihood of the first cluster is the probability of remaining active within the cluster after the first and second visits, multiplied by the probability of leaving the cluster while she is alive after the third visit. Note that, conditional on Γ_i , the values of K_{ij} become fully available to specify p_{ij} in equation (3.6).

Using the general notation of Γ_i in equation (3.9), the probability of the formation of visit clusters Γ_i for customer i is given by:

$$\begin{aligned}
P(\Gamma_i) = & \prod_{c=1}^{m_i-1} \left\{ \prod_{j=u_{ic}}^{v_{ic}-1} (1 - q_i)p_{ij} \right\} (1 - q_i)(1 - p_{iv_{ic}}) \\
& \cdot \prod_{j=u_{im_i}}^{J_i-1} (1 - q_i)p_{ij} \cdot \left\{ (1 - q_i)^{I(v_{im_i} \neq E1)} (1 - p_{iJ_i})^{I(v_{im_i} = E2)} p_{iJ_i}^{I(v_{im_i} = E3)} + q_i^{I(v_{im_i} = E1)} \right\}, \tag{3.11}
\end{aligned}$$

where the first-line expression accounts for the likelihood of the first $(m_i - 1)$ clusters, and the second-line expression accounts for the likelihood of the last cluster and customer defections based on the three different cases.

The overall likelihood function of customer i 's visit patterns can be computed by taking the weighted average of the likelihood function of the observed intervisit times (and the right-censored time periods) associated with each possible formation of visit clusters, where the weights are the probabilities of cluster formation in equation (3.11). Hence, our next step is to derive the likelihood function of the observed intervisit times, conditional on Γ_i . The conditional likelihood function is the product of the density and survival functions in equation (3.4), for which the visit rates of the exponential distributions are determined by Γ_i . In the example illustrated in Figure 3.1, the conditional likelihood function is given by:

$$\begin{aligned}
L(\mathbf{T}_i|\Gamma_i) = & \underbrace{\overbrace{f(t_{i2}|t_{i1}; \lambda_i^w)}^{\text{1st int. time}} \overbrace{f(t_{i3}|t_{i2}; \lambda_i^w)}^{\text{2nd int. time}} \overbrace{f(t_{i4}|t_{i3}; \lambda_i^b)}^{\text{3rd int. time}}}_{\text{1st cluster}} \\
& \cdot \underbrace{f(t_{i5}|t_{i4}; \lambda_i^w) f(t_{i6}|t_{i5}; \lambda_i^w) f(t_{i7}|t_{i6}; \lambda_i^w) f(t_{i8}|t_{i7}; \lambda_i^w) f(t_{i9}|t_{i8}; \lambda_i^b)}_{\text{2nd cluster}} \\
& \cdot \underbrace{f(t_{i10}|t_{i9}; \lambda_i^w)}_{\text{3rd cluster}} S(T|t_{i10}; \lambda_i^b), \tag{3.12}
\end{aligned}$$

where, for example, the first two terms in the first-line expression account for the likelihood of the first and second intervisit times governed by λ_i^w within the first visit cluster, and the last term in the first-line expression accounts for the likelihood of the third intervisit time governed by λ_i^b between the first and second clusters. The survival function in the last line accounts for the right-censored observation.

Using the general notation Γ_i in equation (3.9), the conditional likelihood function of customer i 's intervisit times is given by:

$$\begin{aligned}
L(\mathbf{T}_i|\mathbf{\Gamma}_i) = & \prod_{c=1}^{m_i-1} \left\{ \prod_{j=u_{ic}}^{v_{ic}-1} f(t_{i,j+1}|t_{ij}; \lambda_i^w) \right\} f(t_{iu_{i,c+1}}|t_{iv_{ic}}; \lambda_i^b) \\
& \cdot \prod_{j=u_{im_i}}^{J_i-1} f(t_{i,j+1}|t_{ij}; \lambda_i^w) \cdot \left\{ S(T|t_{iJ_i}; \lambda_i^b)^{I(v_{im_i}=E2)} S(T|t_{iJ_i}; \lambda_i^w)^{I(v_{im_i}=E3)} \right\}, \quad (3.13)
\end{aligned}$$

where the first-line expression accounts for the likelihood of the intervisit times in the first $(m_i - 1)$ visit clusters and the second-line expression accounts for the likelihood of the intervisit times in the last cluster and the right-censored time period.

As discussed, the likelihood function in equation (3.13) is conditional on $\mathbf{\Gamma}_i$, and equation (3.11) is the probability that the cluster formation is given by $\mathbf{\Gamma}_i$. To derive the unconditional likelihood function for customer i , we weight equation (3.13) by equation (3.11) for all possible $\mathbf{\Gamma}_i$'s:

$$L(\mathbf{T}_i) = \sum L(\mathbf{T}_i|\mathbf{\Gamma}_i)P(\mathbf{\Gamma}_i), \quad (3.14)$$

where the summation is over the $2^{J_i-1} \cdot 3$ possible ways of clustering the customer's visits with consideration of the unobserved dropouts. Because the likelihood equation (3.14) is for customer i , the overall likelihood function, after incorporating heterogeneity across N customers, is given by:

$$L_{\text{visit}} = \prod_{i=1}^N \left[\int L(\mathbf{T}_i) dF(\lambda_i^b, \delta_i, \alpha_{0i}, \omega_i) \right]. \quad (3.15)$$

Discussion of the Model

The proposed model extends existing multi-event timing models in a novel way to capture the clustered visit patterns, an important and salient aspect of online

shopping behavior. We compare and contrast our model with several benchmark models to highlight its key properties.

In its basic framework, our visit model is built on the BG/NBD model, an established “buy ‘til you die” model proposed by Fader, Hardie, and Lee (2005). We extend it by relaxing the assumption of a stationary exponential process which postulates that observed intervisit times are drawn from a probability density with a time-invariant visit rate. When there is no consideration of clustered visit patterns (i.e., the probability $p_{ij} = 0$ for all j), the essence of our model reduces to the BG/NBD model.

As we model customers’ switching behavior between the two states which determine the visit rates while allowing for unobserved customer defections, our model is comparable to a hidden Markov model (HMM). HMMs have been applied to model a wide range of latent changes in consumer behaviors (e.g., Montgomery et al. 2004, Netzer, Lattin, and Srinivasan 2008). In its standard form, a HMM is a finite-state stochastic model in which the state of the system at time t (s_t) is not directly observed. Instead, we observe the behavioral outcomes at time t (y_t ; intervisit times in our context), which depends on s_t through the density function $f(y_t|s_t)$. Importantly, a HMM assumes that the transitions between the latent states occur according to the Markov property. In our model, by contrast, the current state depends on states multiple unknown periods back and the order of dependence varies over time, because the probability p_{ij} changes dynamically through customer i ’s visits within a latent visit cluster with unobserved beginning and ending points. Thus, the Markov assumption is flexibly relaxed.

From a modeling standpoint, our model shares similarities with the dynamic

changepoint model proposed by Fader, Hardie, and Huang (2004) for new product sales forecasting. Both models allow a customer's arrival rates to be updated autonomously at any time after a visit (or purchase) to capture underlying non-stationarity in the customer arrival processes. However, a significant difference pertains to how the models specify the probability of changes in visit rates, a key component of the changepoint processes. Fader, Hardie, and Huang (2004) model the probability that a customer updates her purchase rate at the aggregate level and assume the probability is monotonically decreasing as the customer gains more experience with a new product.³ In comparison, we allow the changepoint probability p_{ij} to increase or decrease over time, depending on the formation of visit clusters at the individual customer level, to account for recurring visit clusters in the lumpy shopping patterns of each customer.⁴ In addition to these differences in the model specification, another important difference of the models involves the estimation approach, which we elaborate on below.

3.2.4 Computational Approach

At a glance, maximizing the likelihood function of our model seems a straightforward numerical task. However, looking into $L(\mathbf{T}_i)$ in equation (3.14), we find that the evaluation of the likelihood function to solve the numerical optimization problem entails the consideration of $2^{J_i-1} \cdot 3$ separate changepoint patterns for each customer. For example, when $J_i = 20$, $L(\mathbf{T}_i)$ consists of 1,572,864 ($= 2^{19} \cdot 3$)

³We acknowledge that this assumption is reasonable for Fader, Hardie, and Huang (2004) in the empirical context of their study; customers would become less likely to change their preferences for a new product as they gain more experience with it.

⁴Related to the specification of the changepoint probability, another notable difference is that Fader, Hardie, and Huang (2004) allow the purchase rate to be updated to any nonnegative value at each changepoint. By contrast, the visit rate in our model switches between two values while the customer is alive.

additive terms. In the real world, it is common to observe online customers making many (e.g., 20 or more) shopping trips to a virtual store over a relatively short period of time. Therefore, with the fourfold integrals to consider the multivariate normal density in equation (3.15), evaluating the overall likelihood function would necessitate millions of computations and maximizing the likelihood function would become computationally infeasible for even average-sized panels, or possibly result in local maximum points depending on the initial parameter values.

Facing a similar problem, Fader, Hardie, and Huang (2004) restricted the number of changepoints per customer to a small value (e.g., 4) in their empirical applications. Given that they aimed to forecast new product sales at the aggregate level (while accounting for customer heterogeneity), the restriction on the number of changepoints did not prevent them from attaining the goal and their model was still able to successfully locate critical changepoints governing the evolving sales patterns. For our case though, the numbers of changepoints and visit clusters are interdependently determined by each other, so limiting the number of visit clusters to a small value would jeopardize our main research objective.

In addition to maximizing the complex likelihood function, another important limitation of the maximum likelihood estimation in our case is that the approach does not allow us to infer \mathbf{K}_i (and thus the formation of visit clusters), which could be useful to examine online customers' shopping behavior. These limitations lead us to consider an alternative estimation approach using Bayesian methods through data augmentation.

At the heart of our Bayesian estimation approach is to treat K_{ij} as a latent

variable and draw its samples during the estimation procedure. Conditional on K_{ij} , the specification of p_{ij} in equation (3.6) is nothing but a typical logistic function. However, because K_{ij} can take any positive integer values, we find it difficult to obtain the sample draws directly. We thereby take an indirect approach to infer K_{ij} as follows. First, we define a latent variable H_{ij} that indicates whether customer i stays within the current cluster after the j th visit or leaves the cluster after the j th visit. Then, by the definition of K_{ij} and H_{ij} , we have:

$$K_{ij} = K_{i,j-1} \times H_{i,j-1} + 1. \quad (3.16)$$

That is, K_{ij} increases when $H_{i,j-1} = 1$ (i.e., within the current visit cluster) and is reset to 1 when $H_{i,j-1} = 0$ (i.e., when a new cluster begins). The relationship of the two variables indicates that the value of K_{ij} can be constructed from $H_{i1}, H_{i2}, \dots, H_{i,j-1}$.

We accordingly discuss how H_{ij} can be drawn via data augmentation (Tanner and Wong 1987). The basic idea is that H_{ij} can be obtained, conditional on $H_{ij'}$ for $j' \neq j$ and other model parameters, through recursive algorithms in the Bayesian framework.⁵ Specifically, by applying Bayes' rule, we can show that, for $j = 1, \dots, J_i - 1$, H_{ij} follows a Bernoulli distribution with the following probability:

$$\frac{f(t_{i,j+1}|t_{ij}; \lambda_i^w)P(\mathbf{H}_{ij}^+|H_{ij} = 1, \mathbf{H}_{ij}^-)P(H_{ij} = 1|\mathbf{H}_{ij}^-)}{f(t_{i,j+1}|t_{ij}; \lambda_i^w)P(\mathbf{H}_{ij}^+|H_{ij} = 1, \mathbf{H}_{ij}^-)P(H_{ij} = 1|\mathbf{H}_{ij}^-) + f(t_{i,j+1}|t_{ij}; \lambda_i^b)P(\mathbf{H}_{ij}^+|H_{ij} = 0, \mathbf{H}_{ij}^-)P(H_{ij} = 0|\mathbf{H}_{ij}^-)},$$

where $\mathbf{H}_{ij}^- = \{H_{i1}, H_{i2}, \dots, H_{i,j-1}\}$, $\mathbf{H}_{ij}^+ = \{H_{i,j+1}, H_{i,j+2}, \dots, H_{iJ_i}\}$, and $P(\mathbf{H}_{ij}^+|H_{ij} = 1, \mathbf{H}_{ij}^-)$ and $P(\mathbf{H}_{ij}^+|H_{ij} = 0, \mathbf{H}_{ij}^-)$ can be computed by $\prod_{j'=j+1}^{J_i} P(H_{ij'}|\mathbf{H}_{ij'}^-)$ from equation (3.6). Similarly, H_{iJ_i} can be drawn from a Bernoulli distribution with the following

⁵A similar but relatively simpler recursive estimation approach is used to estimate HMM; see Scott (2002).

probability:

$$\frac{S(T|t_{iJ_i}; \lambda_i^w)P(H_{iJ_i} = 1|\mathbf{H}_{iJ_i}^-)}{S(T|t_{iJ_i}; \lambda_i^w)P(H_{iJ_i} = 1|\mathbf{H}_{iJ_i}^-) + S(T|t_{iJ_i}; \lambda_i^b)P(H_{iJ_i} = 0|\mathbf{H}_{iJ_i}^-)},$$

where we use the survival functions instead of the density functions because the last observation for a customer is right-censored.

To facilitate our Bayesian estimation procedure, we define another latent variable D_{ij} that indicates whether customer i defects permanently after the j th visit. Then, by the definition of q_i , $D_{ij} = 1$ with probability q_i and $D_{ij} = 0$ with probability $(1 - q_i)$. D_{ij} can be sampled as follows. For $j = 1, 2, \dots, J_i - 1$, we know $D_{ij} = 0$ because customer i makes the $(j + 1)$ th visit. By applying Bayes' rule, we can draw D_{iJ_i} from a Bernoulli distribution with probability:

$$\frac{P(D_{iJ_i} = 1)}{S(T|t_{iJ_i}; \lambda_i^w)^{I[H_{iJ_i}=1]} S(T|t_{iJ_i}; \lambda_i^b)^{I[H_{iJ_i}=0]} P(D_{iJ_i} = 0) + P(D_{iJ_i} = 1)}.$$

Once we draw samples for the augmented variables H_{ij} (and thus K_{ij}) and D_{ij} at each iteration of the estimation algorithm, other parameters of the visit and purchase models can be drawn from their respective full conditional distributions using standard Bayesian theory (e.g., Rossi, Allenby, and McCulloch 2006).

To complete the Bayesian specification of the model, we assign priors to the model parameters. Because we lack any prior information, we take the usual route and assign noninformative conjugate priors to the parameters. For each mean parameter, we use a normal density prior. We assume inverse-gamma priors for the variance parameters.

Inferences were obtained using a data augmentation MCMC sampler, implemented in the freely available software, WinBUGS. Results reported are summa-

alized from the output of three independent MCMC chains run for 40,000 iterations, each started from hyper-dispersed starting values, with a burn-in period of 20,000 iterations and utilizing the 60,000 draws (20,000 per chain) thereafter. Convergence was diagnosed both graphically and using the *R*-statistic diagnostic of Gelman and Rubin (1992). We believe that it is of importance to note that the model can be fit in WinBUGS (with ease) once the data augmentation scheme is utilized.

3.3 Data

We use Internet clickstream data collected by comScore from Wharton Research Data Services (wrds.wharton.upenn.edu). The comScore database captures detailed visit and purchase behavior by Internet users across the United States over time. The panel is based on a random sample of users who agreed to install unobtrusive software on their personal computers that monitored their browsing activities. The collected data include the precise day and time when panelists viewed a specific URL at the session level. This allows us to compute the precise intervisit times between online visits (sessions) by individual panelists. The data also contain information about how many webpages the panelists viewed and whether they made a purchase during each visit. This comprehensive database of online clickstream logs has been employed in a number of empirical studies in the marketing literature (e.g., Danaher and Smith 2011, Moe and Fader 2004, Park and Fader 2004).

For our purpose, we use data pertaining to VictoriasSecret.com, an online retail store which sells a variety of women's clothes including lingerie and sleep-

wears. Our data span a period of one year from January 2009 to December 2009. We use the first nine-month data for model calibration and the remaining three-month data for model validation. We sample 1,245 customers who made more than four visits to the online store during the calibration period. The total number of visits made by the customers during the entire data period is 15,041. Out of the total visits, 6.09% (916) resulted in purchase conversions. On average, customers spent 12.2 minutes and viewed 21.9 webpages per visit. In Table 3.1, we provide the summary statistics of the calibration data.

Table 3.1: Descriptive Statistics of the Data

	Mean	Std. Dev.	Min	Max
Observation duration (days)	321.74	51.02	119.21	364.95
Intervisit time (days)	14.09	24.73	0.01	257.47
No. of visits	12.08	13.22	5.00	169.00
No. of purchases	0.74	1.26	0.00	16.00
No. of pageviews per visit	21.93	30.83	1.00	316.00
Time spent per visit (minutes)	12.17	18.52	0.00	247.00

Our data also contain information about websites that referred customers to VictoriasSecret.com. Out of 15,041 visits during the calibration data period, 8,761 visits to the online store were made by directly typing its URL, www.VictoriasSecret.com, in a Web browser, and the remaining 6,280 visits were made through 165 different reference sites by clicking advertising links or keyword search results. Table 3.2 summarizes the frequency of reference sites chosen by customers and the conversion rates at VictoriasSecret.com when customer visits were made through the reference sites. The conversion rate is 0.069 when customers visited the online store by typing its URL (i.e., no reference sites employed). Among the reference sites used by customers, Yahoo.com con-

veyed the largest number of visits to the online retailer and the conversion rate was highest when customer visits were made through Google.com.

Table 3.2: Data Summary on Reference Sites

	Frequency	Conversion rate
No reference sites	8,761	0.069
Yahoo.com	2,355	0.049
Google.com	1,636	0.062
Live.com	799	0.040
AOL.com	311	0.051
Other 161 sites	1,179	0.034

3.4 An Empirical Application

In this section, we explore the empirical performance of our proposed model. We begin by describing the covariates considered in the model. We next discuss the fit of the proposed model in comparison to other benchmark models. Finally, we provide the results based on the estimates of model parameters.

3.4.1 Model Specification

We describe covariates used in our empirical analysis. The vectors of individual-specific time-varying variables, \mathbf{X}_{ij} in equation (3.2) and \mathbf{Z}_{ij} in equa-

tion (3.6), are defined as follows:⁶

- $\mathbf{X}_{ij} = \{\text{PAGEVIEWS}, \text{LPURCHASE}, \text{LINTTIME}, \text{WEEKEND}, \text{HOLIDAY}, \text{YAHOO}, \text{GOOGLE}, \text{LIVE}, \text{AOL}, \text{OTHERS}\}$ and
- $\mathbf{Z}_{ij} = \{\text{PAGEVIEWS}, \text{PURCHASE}, \text{LINTTIME}, \text{WEEKEND}, \text{HOLIDAY}\},$

where PAGEVIEWS is the logarithm of the number of webpages viewed by customer i at her j th visit to the online store; LPURCHASE is a dummy variable indicating whether customer i makes a purchase at her $(j - 1)$ th visit; PURCHASE is a dummy variable indicating whether customer i makes a purchase at her j th visit; LINTTIME is the logarithm of the interval time between customer i 's $(j - 1)$ th and j th visits; WEEKEND is a dummy variable indicating whether customer i 's j th visit is made during the weekend; HOLIDAY is a dummy variable indicating whether customer i 's j th visit is made on days between a national holiday and the weekend ahead of the holiday⁷; {YAHOO, GOOGLE, LIVE, AOL, OTHERS} is a set of dummy variables to consider the reference site employed by customer i for her j th visit to the online store. The baseline condition is when the visit is made by directly typing the URL of the online store (www.victoriassecret.com).

We next define the functional forms of $f(K_{ij})$ in equation (3.2) and $g(K_{ij})$ in equation (3.6). As discussed, the functions capture the customer dynamics within a visit cluster and their effects on the purchase probability and visit patterns, and are defined as follows:

$$f(K_{ij}) = \theta_1 K_{ij} + \theta_2 K_{ij}^2 \quad \text{and} \quad g(K_{ij}) = \psi_1 K_{ij} + \psi_2 K_{ij}^2, \quad (3.17)$$

⁶The duration of a visit is available in the data but not included in the covariate vectors because of its high correlation with the number of pageviews.

⁷The list of national holidays in 2009 obtained from https://www.opm.gov/operating_status_schedules/fedhol/2009.asp.

where we include the square terms of K_{ij} to consider its probable nonlinear effect.⁸

3.4.2 Model Fit

Before presenting detailed results for our proposed model, we demonstrate its superiority in comparison to a number of interesting benchmark models. Our model nests several established models as special cases by modifying behavioral assumptions regarding the formation of latent visit clusters in customers' visit processes. As a restricted version of our model, the first benchmark model (Model 1) employs the Bayesian variant of BG/NBD model described in section 3.2.3 and characterizes the customer visit patterns as a stationary exponential timing process. Note that K_{ij} is not defined in the model because the clustered patterns of customer visits are not considered. In Model 1, we therefore capture customers' purchase decisions using equation (3.2) without the function $f(K_{ij})$.

Model 2 is built on a HMM with three states labeled "alive within a cluster," "alive between clusters," and "dead." The three-state HMM extends the BG/NBD model by allowing for different arrival rates depending on customers' latent states. This model can compute the values of K_{ij} based on customers' switches between states and thus allows us to use the proposed purchase model. However, compared to our visit model, the HMM does not consider the probable presence of latent visit clusters with higher-order dependence across visits and fails to account for ongoing customer dynamics within a visit cluster.

Another benchmark model we consider is constructed based on the notion

⁸Higher-order polynomials of K_{ij} were also estimated and found to be statistically insignificant.

that a purchase conversion may serve as an indicator of the formation of visit clusters. The model assumes that a visit cluster begins right after a purchase and ends with a subsequent purchase. Thus, a visit cluster spans all visits between two consecutive purchase events. We refer to this alternative model as Model 3. Lastly, Model 4 is our proposed model.

To provide an overall comparison of these alternative models, we compute the log marginal density. As a widely used criterion for model comparison in a Bayesian framework, a larger value of the log marginal density indicates a better fit. For another measure of model fit, we compute the hit rate, which refers to the percentage of times that the model correctly predicts a customer's purchase within a next day given her prior visits to the online store.

Table 3.3 shows the model-fit results for all four models. As shown in the table, Model 4, our proposed model, performs best according to both model-fit measures. The better fit of Model 4 compared with Model 1 indicates that the customer visit process is not stationary and thus it is important to take into account lumpy visit patterns when examining online shopping behavior. Model 4 also outperforms Model 2, which suggests the benefits of relaxing the first-order Markov assumption and considering customer dynamics within a visit cluster. Finally, the better fit of Model 4 over Model 3 implies that there are visit clusters that are not associated with purchases. This result highlights the efficacy of our stochastic approach of modeling online shopping patterns, compared with the deterministic approach that postulates all visits between two purchase events consist in the same cluster.

Table 3.3: Model Comparisons

	Log marginal density	In-sample Hit rate	Out-of-sample Hit rate
Model 1	-49,083.5	0.68	0.63
Model 2	-45,603.7	0.76	0.69
Model 3	-48,790.3	0.71	0.64
Model 4	-45,441.0	0.79	0.73

3.4.3 Model Results

We describe inferences based on the estimates of model parameters, obtained by looking at their posterior distributions.

Purchase Model

Table 3.4 reports a summary of the posterior means and their 95% posterior intervals for the parameter estimates of the purchase model. The major findings are as follows. First, we find that the coefficient for PAGEVIEWS is positively significant. This implies that customers are more likely to make a purchase as they view more webpages during a visit. Second, LPURCHASE has a negatively significant effect on the purchase probability. Thus, a customer who made a purchase at her prior visit is less likely to buy at the current visit. Third, we find all covariates constructed based on the timing of visits (i.e., LINTTIME, WEEKEND, HOLIDAY) do not play any significant role in predicting customers' purchase decisions. Fourth, when it comes to the reference sites, we find that customers who visit the online store through YAHOO and other reference sites are less likely to make a purchase compared to those who visit the online store by

directly typing the URL of the retailer, or employ GOOGLE, LIVE, or AOL as a reference site. Lastly, we note that both θ_1 and θ_2 , coefficients for K_{ij} and K_{ij}^2 in the function $f(K_{ij})$, are significant. This implies that a customer's purchase likelihood at a given visit varies depending on how many visits the customer has made previously within a visit cluster. In particular, we find that the effect of K_{ij} on the conversion probability is nonlinear: on average, the conversion probability increases up to the third visit within a cluster and then gradually decreases afterward. However, it is important to note that these results are aggregated across clusters of different sizes.

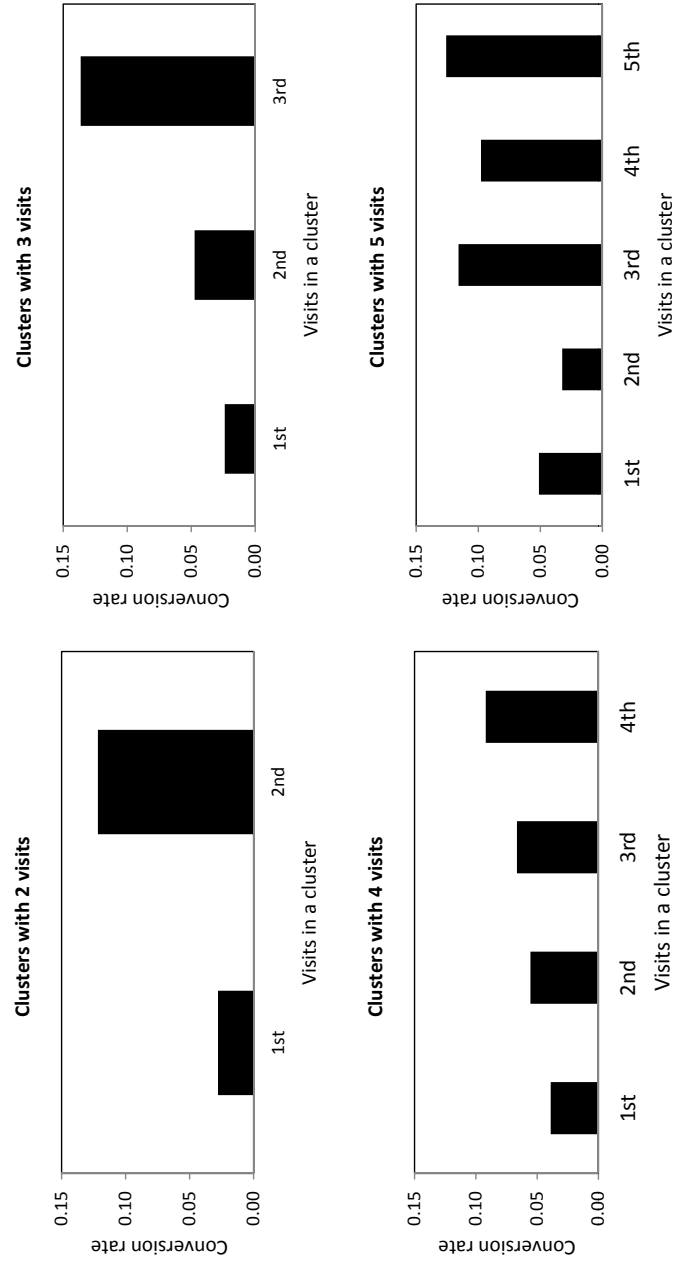
Table 3.4: Parameter Estimates of the Purchase Model

Parameter	Posterior mean	95% Posterior interval
μ_α	-6.36	[-6.61,-6.09]
σ_α	1.44	[1.29, 1.58]
PAGEVIEWS	1.77	[1.66, 1.87]
LPURCHASE	-0.25	[-0.46,-0.02]
LINTTIME	-0.02	[-0.09, 0.04]
WEEKEND	0.10	[-0.10, 0.31]
HOLIDAY	0.34	[-0.12, 0.83]
YAHOO	-0.31	[-0.60,-0.02]
GOOGLE	-0.25	[-0.59, 0.06]
LIVE	-0.33	[-0.79, 0.15]
AOL	-0.24	[-0.81, 0.52]
OTHERS	-0.36	[-0.64,-0.04]
θ_1	0.05	[0.02, 0.09]
θ_2	-0.01	[-0.02,-0.01]

To further examine the likelihood of purchase conversion, we compute the conversion rates across customer visits within a cluster of same size. This result can be obtained once we infer the formation of visit clusters using the estimates

of the augmented variable K_{ij} . We find that conversion rate varies substantially depending on the size of a visit cluster and the location of a visit in the cluster. Specifically, when a cluster consists of only one visit, the conversion rate at the visit is 0.05. Figure 3.2 shows how the conversion rate changes within a cluster when the cluster consists of more than one visit. For clusters with two visits, the conversion rate ranges from 0.03 at the first visit to 0.12 at the second visit of the visit cluster. For clusters with three visits, the conversion rate is 0.02 at the first visit, 0.05 at the second visit, and 0.14 at the third visit of the cluster. As illustrated in the figure, a considerable portion of the purchase conversions occur at later visits of the cluster, and the likelihood of making purchases appears to be higher at the last visit of the cluster.

Figure 3.2: Conversion Rate in a Visit Cluster



As another way of demonstrating the benefit of clustering customer visits, we combine the inferred formation of visit clusters with the data of the reference sites for VictoriasSecret.com, and compute the conversion rates for each sequence of reference sites chosen by customers within a cluster. Table 3.5 reports the five most frequently chosen sequences of reference sites and the corresponding conversion rates for clusters with two visits. The table shows that the conversion rate varies considerably depending on the sequence of reference sites chosen by customers. In particular, when customers made their both first and second visit of the visit cluster through YAHOO, the conversion rates at the first and second visits were 0.02 and 0.09, respectively. In sharp contrast, when customers made their first visit through YAHOO and the second visit by directly typing the URL of the online store, they made purchase with a probability of 0.02 at the first visit and with a probability of 0.15 at the second visit of the visit cluster. In general, we find that, when a customer's first visit was made through any of reference sites, the conversion rate at her second visit was much higher when the visit was directly made by customers typing the URL of the online store, compared to the case where the second visit was also made through reference sites. We also find similar patterns of conversion rates for clusters with more than two visits. These results suggest that the cluster-based analysis of customer visits can help marketers understand and predict online purchase behavior across a series of store visits.

Table 3.5: Conversion Rates for the Sequence of Reference Sites in Clusters with Two Visits

			Conversion rate	
			At the first visit	At the second visit
No reference site	→	No reference site	0.03	0.14
Yahoo.com	→	No reference site	0.02	0.15
Yahoo.com	→	Yahoo.com	0.02	0.09
Google.com	→	No Reference site	0.03	0.18
Google.com	→	Google.com	0.02	0.11

Visit Model

We next discuss the results regarding the underlying customer visit patterns. Tables 3.6 and 3.7 report a summary of the posterior means and their 95% posterior intervals for the parameter estimates of the visit model. From the estimates of μ_λ , μ_δ , and Σ , we find that the mean visit rate between clusters (λ_i^b) is 0.04.⁹ Therefore, the mean intervisit time between clusters is about 25 ($=1/0.04$) days. In contrast, the mean visit rate within a cluster (λ_i^w) is 17.68, implying the mean intervisit time within a cluster is 0.06 days (about 1.4 hours). Customers thus make considerably more frequent visits within a cluster. In comparison, we find that the mean visit rate estimated under Model 1 is 0.07, which can be converted into a mean intervisit time of 14 days. Note that the mean visit rate under Model 1 falls between the two mean visit rates (λ_i^w and λ_i^b) under our proposed model. The estimate of μ_ω indicates that on average, customers defect with a probability of 0.03 after their each visit to the online store.

Our next inferences are based on the estimates of covariates which govern the probability p_{ij} and thus characterize the formation of latent visit clusters in the visit patterns. The major findings are as follows. First, we find that the coefficient for PAGEVIEWS is positively significant. Thus, as a customer views more webpages during a visit, she is more likely to make another visit within the current cluster. Second, a negatively significant coefficient for PURCHASE suggests that if a customer makes a purchase, she tends to leave the visit cluster. However, LINTTIME, WEEKEND, and HOLIDAY have no significant impact on p_{ij} . Importantly, we find that both ψ_1 and ψ_2 , coefficients for K_{ij} and K_{ij}^2 in the function $g(K_{ij})$, are significant. Thus, a customer's likelihood of making

⁹Note that if $X \sim N(\mu, \sigma^2)$, $Y = e^X$ follows a log-normal distribution with a mean of $e^{\mu+\sigma^2/2}$.

Table 3.6: Parameter Estimates of the Visit Model

Parameter	Posterior mean	95% Posterior interval
μ_λ	-3.51	[-3.57,-3.45]
μ_δ	0.61	[0.45, 0.76]
μ_β	-1.59	[-1.78,-1.42]
μ_ω	-3.46	[-3.65,-3.26]
PAGEVIEWS	0.26	[0.22, 0.30]
PURCHASE	-1.01	[-1.24,-0.77]
LINTTIME	-0.09	[-0.19, 0.02]
WEEKEND	-0.05	[-0.18, 0.08]
HOLIDAY	0.11	[-0.14, 0.38]
ψ_1	0.37	[0.32, 0.43]
ψ_2	-0.01	[-0.02,-0.01]

Table 3.7: Estimated Σ

	$\log \lambda_i^b$	$\log \delta_i$	β_{0i}	ω_i
$\log \lambda_i^b$	0.54 [0.46, 0.63]	0.85 [0.70, 1.02]	-0.35 [-0.45,-0.26]	0.16 [0.11, 0.22]
$\log \delta_i$		4.52 [4.08, 5.05]	-1.90 [-2.22,-1.61]	1.37 [1.21, 1.57]
β_{0i}			0.87 [0.66, 1.07]	-0.60 [-0.71,-0.50]
ω_i				0.47 [0.38, 0.57]

clustered visits dynamically varies depending on how many visits the customer has made previously within a visit cluster.

Finally, the variance estimates in Table 3.7 show that customers are heterogeneous in their lumpy visit patterns. The significant covariance estimates sug-

gest that it is important to consider interdependence between the model components.

One of the key benefits of our Bayesian estimation approach is that we can obtain the estimates of customer-specific parameters by treating them as model parameters, rather than integrating them out as in the maximum likelihood approach. Thus, we can employ the estimates at the individual level for customization purposes. In particular, through the augmented variable K_{ij} , our model provides a set of novel inferences about the store visit patterns by customers.

Table 3.8 reports the posterior means of several customer-specific statistics, computed from the estimates of individual-specific parameters across the iterations of the MCMC samplers: (1) the intervisit time within a cluster, (2) the number of visits per cluster, (3) the time length of a cluster, (4) the intervisit time between clusters, and (5) the number of visit clusters during the calibration period, for customers at every 10th percentile with respect to their total number of visits during the period. The results show that the visit patterns by the customers are very lumpy, and the extent of the lumpiness differs considerably across customers. The intervisit time within a cluster ranges from 0.08 days (about 2 hours) to 1.58 days while the intervisit time between clusters ranges from 2.28 to 72.94 days. The mean number of visits per cluster ranges from 1.2 to 3.9 across the customers. Using the intervisit time within a cluster and the number of visits per cluster, we compute that a visit cluster spans less than a day for some customers and a few days for others.

Table 3.8: Customer-Specific Statistics of Latent Visit Clusters

Decile	Total No. of visits	Intervisit time within a cluster	No. of visits per cluster	Time length of a cluster	Intervisit time between clusters	No. of clusters
Max	169	0.08	1.39	0.03	2.28	121.23
10%	23	1.26	2.93	2.44	19.03	7.84
20%	15	0.12	1.45	0.06	15.63	10.34
30%	12	0.14	3.94	0.42	10.63	3.04
40%	10	1.58	1.87	1.38	36.36	5.35
Median	8	0.33	1.39	0.13	20.96	5.76
60%	7	0.52	2.71	0.89	55.13	2.58
70%	6	0.68	1.51	0.35	43.10	3.97
80%	6	0.26	1.20	0.05	32.78	5.79
90%	5	0.22	1.21	0.05	51.15	4.13
Min	5	0.78	2.42	1.12	72.94	2.06

The comparisons between individual customers further highlight the managerial benefits of our model in targeting customers with timely marketing actions. Table 3.9 reports the cluster-based statistics for two customers who both made 37 visits to the online store during the calibration period. The customers could be treated equally under the extant multi-event timing models if their recencies are similar to each other. In contrast, our results indicate that their underlying store visit patterns differ considerably. Customer 51 made much less visits within a cluster than customer 52 did (i.e., 1.27 versus 3.80) but formed much more visit clusters in the visit process (i.e., 29.11 versus 6.73). The two customers also greatly differ in their intervisit times. Customer 51's average intervisit time within a cluster is 0.08 days and that between clusters is 10.04 days. In contrast, customer 52's average intervisit time within a cluster is 1.13 days and that between clusters is 27.43 days.

These results suggest that the online retailer take different marketing approaches for these two customers. Specifically, customer 52 tends to make multiple store visits once she is in an active state of shopping, which normally lasts for several days. When the customer's first or second visits are ended up with no purchases, the online retailer could benefit by following up her browsing behavior during the earlier visits within a cluster, and being prepared with customized offers and services for upcoming sales opportunities. On the other hand, because customer 51 is more likely to make only a couple of clustered visits in a relatively short period of time, the online retailer may consider taking preemptive marketing actions based on her purchase history rather than the "wait-and-see" approach taken for customer 52.

Table 3.9: Comparison of Two Customers

Customer ID	Total No. of visits	Intervisit time within a cluster	No. of visits per cluster	Time length of a cluster	Intervisit time between clusters	No. of clusters
⋮	⋮	⋮	⋮	⋮	⋮	⋮
51	37	0.08	1.27	0.02	10.04	29.11
52	37	1.13	3.80	3.16	27.43	6.73
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Finally, we note that the inferences of latent visit clusters can help marketers conduct attribution management online. With low conversion rates, online retailers have sought to understand the underlying relationship between a series of customer visits and purchase conversions (*Forrester Research* 2009). Given that visit clusters could serve as a reasonable proxy for business opportunities with customers, the cluster formation available from our model enables marketers to infer which customer visits lead to sales (or have more influences on sales) by linking a purchase conversion with store visits occurred in the same cluster. In turn, this allows marketers to evaluate the effectiveness of their online initiatives and campaigns, and assign more resources to successful marketing actions associated with influential visits.

3.5 Conclusions and Future Research

We develop a dynamic model that captures the lumpy visit patterns of online customers and predicts purchase conversions across store visits. A major benefit of this model is its ability to infer the formation of latent visit clusters in the arrival process of customer visits, which offers a better understanding of online shopping behavior.

Our approach to modeling the lumpy shopping patterns at the individual customer level assumes that the arrival process of customer visits follows a mixture of two timing processes, where the weight for each process changes dynamically as a customer progresses through multiple visits in a latent visit cluster. Because the start and the end of each visit cluster are not observed, we employ a changepoint modeling framework and statistically infer the formation of visit

clusters on the basis of customer visit patterns through data augmentation in Bayesian approach. The proposed model also accounts for customer defections and various sources of customer heterogeneity in a flexible manner.

Using Internet clickstream data from VictoriasSecret.com, we demonstrate that the proposed model exhibits excellent fit and predictive performance in a comparison with existing multi-event timing models. Our dynamic model of customer visits also offers several novel inferences about the patterns that underlie online customer shopping behavior such as (1) the intervisit time within a cluster, (2) the number of visits per cluster, (3) the time length of a cluster, (4) the intervisit time between clusters, and (5) the number of visit clusters in a given period, at the individual customer level.

Given the formation of latent visit clusters inferred from the model, we examine online customers' purchase behavior across store visits. We find that the likelihood of purchase conversions varies depending on not only the cumulative number of visits within a cluster but also the sequence of reference sites chosen to visit the online store. These results, together with the results of the visit model, could help marketing managers take timely marketing actions through a better prediction of customer visit patterns and purchase conversions at a given visit.

There are several limitations that should be acknowledged and perhaps addressed in future research. First, one could extend our modeling framework by allowing for other sources of nonstationarity, in addition to the lumpiness in customer shopping patterns. Fader, Hardie, and Huang (2004) show that purchasing behavior for a new product evolves over time. Schweidel and Fader (2009) also report empirical evidence of an evolving process of customer con-

sumption patterns. It would be fruitful to integrate other sources of nonstationarity into our proposed model of customer visits and disentangle their effects on observed behavior.

Second, this research has focused on studying customer shopping behavior for a single online retailer. In the online world, customers often visit multiple websites to accumulate information and compare several retailers before making a purchase decision. For example, Neslin et al. (2006) and Neslin and Shankar (2009) suggest the importance of multichannel customer management. Park and Fader (2004) show that combining cross-site browsing patterns for two websites explains customer behavior better in both sites. Along similar lines, restricting the analysis to a single site can leave out other important touchpoints that lead to a conversion. Another area for future research therefore is to investigate the competing nature of multiple websites and their role in customers' purchase behavior. We hope this study generates further interest in exploring various underlying patterns in customer data, which can help us develop deeper insights into customer behavior.

References

- Barry, Daniel and John A. Hartigan (1993), "A Bayesian Analysis for Change-point Problems," *Journal of the American Statistical Association*, 88 (421), 309-319.
- Chintagunta, Pradeep K. and Sudeep Haldar (1998), "Investigating Purchase Timing Behavior in Two Related Product Categories," *Journal of Marketing Research*, 35 (1), 43-53.
- Dacey, Michael F. and Tze-hsiung Tung (1962), "The Identification of Randomness in Point Patterns," *Journal of Regional Science*, 4 (1), 83-96.
- Danaher, Peter J. and Michael S. Smith (2011), "Modeling Multivariate Distributions Using Copulas: Applications in Marketing," *Marketing Science*, 30 (1), 4-21.
- Fader, Peter S., Bruce G. S. Hardie, and Chun-Yao Huang (2004), "A Dynamic Change-point Model for New Product Sales Forecasting," *Marketing Science*, 23 (3), 280-303.
- Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005), "Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model," *Marketing Science*, 24 (2), 275-284.
- Forrester Research (2009), "Interactive Attribution," October 20.
- Forrester Research (2010), "US Online Retail Forecast, 2009 To 2014," March 5.
- Gelman, Andrew and Donald B. Rubin (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7 (4), 457-472.
- Gupta, Sunil (1991), "Stochastic Models of Interpurchase Time with Time-Dependent Covariates," *Journal of Marketing Research*, 28 (1), 1-15.
- Hauser, John R., Glen L. Urban, Guilherme Liberali, and Michael Braun (2009), "Website Morphing," *Marketing Science*, 28 (2), 202-223.
- Jain, Dipak C. and Naufel J. Vilcassim (1991), "Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach," *Marketing Science*, 10 (1), 1-23.
- Johnson, Eric J., Wendy W. Moe, Peter S. Fader, Steven Bellman, and Jerry Lohse (2004), "On the Depth and Dynamics of World Wide Web Shopping Behavior," *Management Science*, 50 (3), 299-308.

- Moe, Wendy W. and Peter S. Fader (2004), "Dynamic Conversion Behavior at E-Commerce Sites," *Management Science*, 50 (3), 326-335.
- Montgomery, Alan L., Shibo Li, Kannan Srinivasan, and John C. Liechty (2004), "Modeling Online Browsing and Path Analysis Using Clickstream Data," *Marketing Science*, 23 (4), 579-595.
- Moorthy, Sridhar, Brian T. Ratchford, and Debabrata Talukdar (1997), "Consumer Information Search Revisited: Theory and Empirical Analysis," *Journal of Consumer Research*, 23 (4), 263-277.
- Morrison, Donald G. and David C. Schmittlein (1981), "Predicting Future Random Events Based on Past Performance," *Management Science*, 27 (9), 1006-1023.
- Morrison, Donald G. and David C. Schmittlein (1988), "Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is It Worth the Effort?," *Journal of Business and Economic Statistics*, 6 (2), 145-159.
- Neslin, Scott A., Dhruv Grewal, Robert Leghorn, Venkatesh Shankar, Marije L. Teerling, Jacquelyn S. Thomas, and Peter C. Verhoef (2006), "Challenges and Opportunities in Multichannel Customer Management," *Journal of Service Research*, 9 (2), 95-112.
- Neslin, Scott A. and Venkatesh Shankar (2009), "Key Issues in Multichannel Customer Management: Current Knowledge and Future Directions," *Journal of Interactive Marketing*, 23 (1), 70-81.
- Netzer, Oded, James M. Lattin, and V. Srinivasan (2008), "A Hidden Markov Model of Customer Relationship Dynamics," *Marketing Science*, 27 (2), 185-204.
- Park, Young-Hoon and Peter S. Fader (2004), "Modeling Browsing Behavior at Multiple Websites," *Marketing Science*, 23 (3), 280-303.
- Pievatolo, Antonio and Renata Rotondi (2000), "Analysing the Interevent Time Distribution to Identify Seismicity Phases: A Bayesian Nonparametric Approach to the Multiple-Changepoint Problem," *Applied Statistics*, 49 (4), 543-562.
- Rossi, Peter E., Greg M. Allenby, and Rob McCulloch (2006), *Bayesian Statistics and Marketing*, New York: John Wiley & Sons.
- Schmittlein, David C., Donald G. Morrison, and Richard Colombo (1987), "Counting Your Customers: Who Are They and What Will They Do Next?," *Management Science*, 33 (1), 1-24.

- Schweidel, David A. and Peter S. Fader (2009), "Dynamic Changepoints Revisited: An Evolving Process Model of New Product Sales," *International Journal of Research in Marketing*, 26 (2), 119-124.
- Schweidel, David A., Peter S. Fader, and Eric T. Bradlow (2008), "A Bivariate Timing Model of Customer Acquisition and Retention," *Marketing Science*, 27 (5), 829-843.
- Scott, Steven L. (2002), "Bayesian Methods for Hidden Markov Models: Recursive Computing in the 21st Century," *Journal of the American Statistical Association*, 97 (457), 337-351.
- Seetharaman, P. B. and Pradeep K. Chintagunta (2003), "The Proportional Hazard Model for Purchase Timing: A Comparison of Alternative Specifications," *Journal of Business and Economic Statistics*, 21 (3), 368-382.
- Singh, Siddharth S., Sharad Borle, and Dipak C. Jain (2009), "A Generalized Framework for Estimating Customer Lifetime Value When Customer Lifetimes are Not Observed," *Quantitative Marketing and Economics*, 7 (2), 181-205.
- Sismeiro, Catarina and Randolph E. Bucklin (2004), "Modeling Purchase Behavior at an E-Commerce Web Site: A Task Completion Approach," *Journal of Marketing Research*, 41 (3), 306-323.
- Tanner, Martin A. and Wing Hung Wong (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82 (398), 528-540.
- Telang, Rahul, Peter Boatwright, and Tridas Mukhopadhyay (2004), "A Mixture Model for Internet Search-Engine Visits," *Journal of Marketing Research*, 41 (2), 206-214.
- Urbany, Joel E., Peter R. Dickson, and William L. Wilkie (1989) "Buyer Uncertainty and Information Search," *Journal of Consumer Research*, 16 (2), 208-215.
- Wedel, Michel, Wagner A. Kamakura, Wayne S. DeSarbo, and Frenkel Ter Hofstede (1995), "Implications for Asymmetry, Nonproportionality and Heterogeneity in Brand Switching from Piece-wise Exponential Mixture Hazard Models," *Journal of Marketing Research*, 32 (4), 457-462.